Regulatory Toxicology and Pharmacology 72 (2015) 514–537

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence ☆

Richard A. Becker [a,*], Gerald T. Ankley [b], Stephen W. Edwards [c], Sean W. Kennedy [d], Igor Linkov [e], Bette Meek [f], Magdalini Sachana [g], Helmut Segner [h], Bart Van Der Burg [i], Daniel L. Villeneuve [b], Haruna Watanabe [j], Tara S. Barton-Maclaren [k]

[a] American Chemistry Council, Washington, DC, USA
[b] US Environmental Protection Agency, Office of Research and Development, Mid-Continent Ecology Division, Duluth, MN, USA
[c] US Environmental Protection Agency, Office of Research and Development, Integrated Systems Toxicology Division, Research Triangle Park, NC, USA
[d] Environment Canada, National Wildlife Research Centre, Ottawa, Ontario, Canada
[e] Environmental Laboratory, Engineer Research and Development Center, US Army Corps of Engineers, Concord, MA, USA
[f] University of Ottawa, Ottawa, Ontario, Canada
[g] European Commission, Joint Research Centre, Via E. Fermi 2749, 21027, Ispra, Italy
[h] Centre for Fish and Wildlife Health, University of Bern, Bern, Switzerland
[i] BioDetection Systems BV, Amsterdam, The Netherlands
[j] Center for Environmental Risk Research, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan
[k] Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 11 March 2015
Available online 8 April 2015

Keywords:
Adverse outcome pathway
Weight of evidence
Mode of action
Bradford-Hill considerations

ABSTRACT

Systematic consideration of scientific support is a critical element in developing and, ultimately, using adverse outcome pathways (AOPs) for various regulatory applications. Though weight of evidence (WoE) analysis has been proposed as a basis for assessment of the maturity and level of confidence in an AOP, methodologies and tools are still being formalized. The Organization for Economic Co-operation and Development (OECD) Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs (OECD 2014a; hereafter referred to as the OECD AOP Handbook) provides tailored Bradford-Hill (BH) considerations for systematic assessment of confidence in a given AOP. These considerations include (1) biological plausibility and (2) empirical support (dose-response, temporality, and incidence) for Key Event Relationships (KERs), and (3) essentiality of key events (KEs). Here, we test the application of these tailored BH considerations and the guidance outlined in the OECD AOP Handbook using a number of case examples to increase experience in more transparently documenting rationales for assigned levels of confidence to KEs and KERs, and to promote consistency in evaluation within and across AOPs. The major lessons learned from experience are documented, and taken together with the case examples, should contribute to better common understanding of the nature and form of documentation required to increase confidence in the application of AOPs for specific uses. Based on the tailored BH considerations and defining questions, a prototype quantitative model for assessing the WoE of an AOP using tools of multi-criteria decision analysis (MCDA) is described. The applicability of the approach is also demonstrated using the case example aromatase inhibition leading to reproductive dysfunction in fish. Following the acquisition of additional experience in the development and assessment of AOPs, further refinement of parameterization of the model through expert elicitation is recommended. Overall, the application of quantitative WoE approaches hold promise to enhance the rigor, transparency and reproducibility for AOP WoE determinations and may play an important role in delineating areas where research would have the greatest impact on improving the overall confidence in the AOP.

☆ This manuscript was reviewed in accordance with guidelines of the authors' employers, but does not necessarily reflect policy/positions of the USEPA, JRC, Health Canada or other authors' organizations.

* Corresponding author at American Chemistry Council, 700 2nd St., NE, Washington, DC 20002, USA. Fax: +202 478 2503.
E-mail address: rick_becker@americanchemistry.com (R.A. Becker).

http://dx.doi.org/10.1016/j.yrtph.2015.04.004
0273-2300/

List of Abbreviations

adverse outcome pathways (AOPs)
weight of evidence (WoE)
Organization for Economic Co-operation and Development (OECD)
modes of action (MoA)
key event (KE)
Key Event Relationship (KER)
molecular initiating event (MIE)
adverse outcome (AO)
(Quantitative) Structure Activity Relationship ((Q)SAR)
Bradford-Hill (BH)
Arylhydrocarbon receptor (AhR)
cytochrome P450 (CYP)
neuropathy target esterase (NTE)
hexavalent chromium (Cr(VI))
testosterone (T)
17β-estradiol (E2)
vitellogenin VTG
messenger RNA (mRNA)
estrogen receptor α (ERα)
estrogen receptor (ER)
high (H)
moderate (M)
low (L)
lines of evidence (LOE)
multi-criteria decision analysis (MCDA)

1. Introduction

A large number of substances in commerce require risk evaluation to protect human health and the environment. A key challenge for the regulatory community is assessing the potential for risks of substances with limited toxicity or toxicology data. Accordingly, various regulatory mandates and related initiatives in Canada, USA, the European Union and, more recently, the Asian Pacific region (see, for example, Council of Labor Affairs, Taiwan, 2012; Dellarco et al., 2010; European Commission, 2006; Hughes et al., 2009; Lowell Center for Sustainable Production, 2012; Meek and Armstrong, 2007, Mitchell et al., 2013) reflect the rapidly growing need for more efficient methods and novel strategies to assess the hazards and risks of a wide array of chemicals. Due to costs and time involved, as well as the desire to reduce animal use in response to ethical considerations, traditional resource-intensive standard *in vivo* toxicology studies are not feasible for the regulatory testing of all chemicals requiring evaluation. Adverse outcome pathways (AOPs) hold great promise as important tools to enhance efficiencies and the future success of risk assessment in the implementation of pathway- and mechanistic-based approaches that are able to accommodate substances and groups of substances with varying amounts and types of toxicological information (e.g., OECD, 2014b; CCA, 2012; Ankley et al., 2010; NRC, 2010; NRC, 2007). However, it is important to note that these promising concepts and approaches supporting the application of pathway-based data and predictive modeling in hazard characterization and risk assessment need further development, evaluation and acceptance before being used routinely.

The generation and consideration of mechanistic data has the potential to increase our understanding of the modes of action (MoA) underlying the toxicity of various individual chemicals and groups of chemicals. It is anticipated that MoA information will lead to improved estimation of potential risk to human health and the environment. Investigators continue to elucidate the modes and mechanisms underlying toxicity-related adverse effects by applying emerging and increasingly more sophisticated computational, molecular and *in vitro* technologies. Such approaches have the potential to be used qualitatively and/or quantitatively in a predictive manner to identify potential toxicities in the absence of definitive data on adverse effects. A major challenge faced by both research and regulatory scientists is the integration of data and information being generated from diverse sources at many different levels of biological organization in a manner that is transparent, informative and suitable for regulatory decision-making.

Conceptually, an AOP is similar to a MoA (OECD, 2013) with the MoA representing a chemical and species specific application of the more general AOP. The AOP construct (Fig. 1) portrays a MoA in a structured framework that organizes and links knowledge of Key Events (KEs; a change in biological state that is both measurable and essential to the progression of a defined biological perturbation) in a sequence that commences with the molecular initiating event (MIE; the initial point of chemical-biological interaction within the organism that starts the pathway) and proceeds through a series of higher order biological events, culminating with the *in vivo* adverse outcome (AO) of interest to risk assessment. The series of biological events, or KEs, are connected to one another via linkages defined as Key Event Relationships (KERs). An AOP that is anchored to both a MIE and an AO provides a consistent structure that facilitates effective application and integration of diverse information on MoAs for various hazard and risk assessment uses, and provides a tool for the identification of key uncertainties and research needs (Ankley et al., 2010; OECD, 2013). Villeneuve et al. (2014a,b) provide detailed discussion of definitions of MIEs, KEs and KERs as well as strategies, principles and best practices to use when developing AOPs, and refer to work reported here with respect to conduct of WoE evaluations; other products from the 2014 workshop "Advancing AOPs for Integrated Toxicology and Regulatory Applications" can be found at https://aopkb.org/saop/workshops/somma.html#manuscripts.

Under the auspices of the Organisation for Economic Cooperation and Development (OECD), scientists across the world and from all sectors have an opportunity to develop AOPs which will be peer- reviewed and publically accessible through a wiki-based tool (AOP-Wiki; aopwiki.org). Using the wiki format, contributions to improving the scientific basis and range of applications of AOPs can be made by experts from all sectors and regions. When fully actualized, the AOP-Wiki will serve as a repository of information of AOPs, KEs and KERs for a wide spectrum of toxicologically-relevant pathways. This organized and integrated information is envisioned to address or inform a number of analytical domains in the decision-making process including: (1) efficient grouping of chemicals based on common pathways of toxicity and potential consideration of non-test methods, such as read-across and (quantitative) structure-activity relationship ((Q)SAR) modeling or targeted testing to fill data needs; (2) identification of research priorities relevant to data gaps in regulatory test batteries; (3) providing a framework for priority setting; and, (4) hazard characterization and risk assessment that incorporate qualitative and quantitative determinations of human and/or ecological relevance and variability, dose–response extrapolation and potential for combined effects of chemicals (OECD, 2013, 2014a,b; Meek et al., 2014a,b).

In order for AOPs to be considered for a specified application by the regulatory community, it is critical to standardize AOP development and provide a clear and transparent evaluation of

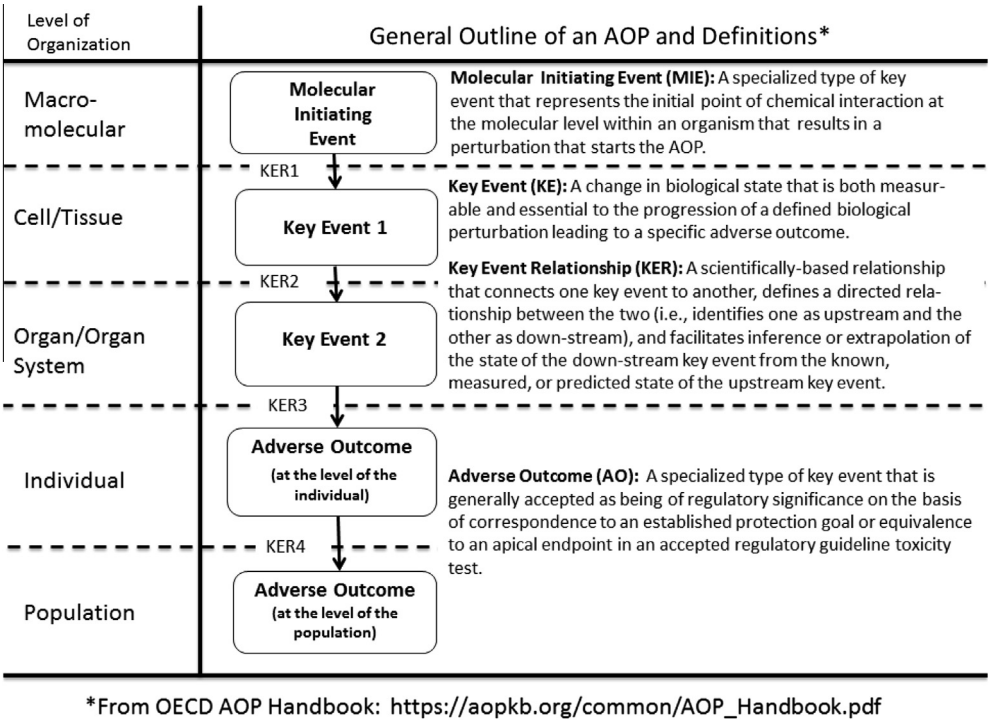

Fig. 1. Framework of an AOP.

reliability, robustness and relevance. Consequently, weight of evidence (WoE) evaluation (Weed, 2005; Linkov et al., 2009, 2015; Rhomberg et al., 2013) was incorporated as a key element of the OECD guidance on developing and assessing AOPs (OECD, 2013). While approaches for conducting WoE evaluations may differ, the essence of all approaches requires considering the collective body of evidence to address the specific questions at hand. The purpose of a WoE evaluation is to document certainty in inferring responses beyond interpolation within the range of empirical observations in a transparent manner. Confidence in inference is underpinned by the degree of certainty that the lines of evidence support the hypothesized inference. The Bradford-Hill (BH) considerations (Hill, 1965), originally developed for the evaluation of causality of associations observed in epidemiological studies, and more recently evolved to increase consistency for WoE determinations in the application of MoA/species concordance analysis (Meek et al., 2014a,b) , provide a useful approach for evaluating the extent of support for hypothesized AOPs. Thus, the BH considerations have been adopted for assessing the WoE of KEs, KERs and overall AOPs (OECD, 2014a). However, there is a need for tailoring of these considerations to address the AOP context (i.e., the non-chemical related elements of the pathway between exposure and effect). Moreover, there is a need for explicit examination and illustration using case examples of how these tailored BH considerations can be applied for the practical purpose of AOP assessment, both qualitatively and, potentially, quantitatively (Linkov et al., 2009, 2015).

Improvement of WoE analysis in the development of AOPs was addressed in a workshop held March 2–7, 2014 in Somma Lombardo, Italy entitled "Advancing AOPs for Integrated Toxicology and Regulatory Applications." The specific objectives of one of the workshop groups were to: (1) Provide guidance and documentation for applying the evolved BH considerations for the WoE evaluation of KEs, KERs and overall AOPs; (2) Illustrate the application of the WoE approach using the evolved BH considerations through the use of case examples; (3) Illustrate how confidence levels in an AOP can help to inform different regulatory applications; (4) Explore challenges and opportunities for constructing a quantitative WoE methodology for AOPs; and, (5) Contribute to the development of a transparent, user friendly approach that has been incorporated in the AOP-Wiki to promote greater consistency in AOP development and facilitate communication of the scientific confidence in an AOP for regulatory applications.

Importantly, preparation of this manuscript has been coordinated with the development of the OECD AOP Handbook (OECD, 2014a) for which there is overlap of contributors. Specifically, the guidance in the OECD AOP Handbook (OECD, 2014a) has been informed and "tested" by example. The case examples and early testing illustrated in this paper are essential in continuing evolution of the guidance, including clarification of the elements of WoE and associated terminology and illustration and revision of categories of the extent of WoE based on practical application. Furthermore, this testing supports advancement in the development of AOPs and provides practical recommendations to facilitate consistency in the evaluation and documentation of WoE of AOPs. Accordingly, lessons learned in the application of the guidance are highlighted. Finally, a prototype multi-criteria decision analysis (MCDA) model is introduced to illustrate an approach that may prove fruitful in the future for quantitatively weighting relevant BH considerations in AOP WoE analyses.

2. Application of evolved BH considerations for WoE in developing an AOP

The challenges in conducting WoE evaluations for diverse AOPs and opportunities for improving the systematic WoE analysis to enhance communication of the scientific confidence in KEs, KERs and AOPs were discussed. The basis for these discussions included the initial evolution of WoE considerations included in a recent update of the human-oriented more chemical specific MoA/species concordance framework (Meek et al., 2014a,b), and the initial tables and narrative on WoE in the OECD guidance (OECD, 2013). Subsequently, in an iterative manner, the

Table 1
Guidance for assessing relative level of confidence in the overall AOP based on evolved Bradford-Hill weight of evidence considerations.[1] [2] [3] [4]

Biological Plausibility *Defining Question: Is there a mechanistic (i.e., structural or functional) relationship between Key Event$_{upstream}$ and Key Even$_{downstream}$ consistent with established biological knowledge?*		
High (Strong) Confidence: Extensive understanding of the Key Event Relationship based on extensive previous documentation and broad acceptance (e.g., mutation leading to tumors), i.e., an established mechanistic basis.	Moderate Confidence: The Key Event Relationship is plausible based on analogy to accepted biological relationships but scientific understanding is not completely established.	Low (Weak) Confidence: There is empirical support for a association between Key Events (see Empirical Evidence below), but the structural or functional relationship between them is not understood.
Essentiality[5] *Defining Question: Are downstream Key Events and/or the Adverse Outcome prevented if an upstream Key Event is blocked?*		
High (Strong) Confidence: Direct evidence from specifically designed experimental studies illustrating essentiality for at least one of the important Key Events (e.g., stop/reversibility/ recovery studies, antagonism, knockout models, etc.)	Moderate Confidence: Indirect evidence that sufficient modification of an expected modulating factor attenuates or augments a Key Event (e.g., augmentation of proliferative response in the Key Event$_{upstream}$ leading to an increase in Key Event$_{downstream}$ or in the Adverse Outcome).	Low (Weak) Confidence: No or contradictory experimental evidence of the essentiality of any of the Key Events.
Empirical Evidence[6] [7] *Defining Questions: Does the empirical evidence support that a change in Key Event$_{upstream}$ leads to an appropriate change in KeyEvent$_{downstream}$? Does KeyEvent$_{upstream}$ occur at lower doses and earlier time points than KeyEvent$_{downstream}$ and is the incidence of KeyEvent$_{upstream}$ greater than that for the KeyEvent$_{downstream}$? Are there inconsistencies in empirical support across taxa, species and stressors that don't align with an expected pattern for the hypothesized AOP? (Note: In many cases, evidence that contributes to quantitative understanding of a Key Event Relationship description will also provide empirical support for the relationship, and such relevant information should be considered as part of the overall weight of evidence evaluation of the concordance of empirical observations and consistency of the Key Event Relationship.)*		
High (Strong) Confidence: Multiple studies showing dependent change in both events following exposure to a wide range of specific stressors. Extensive evidence for temporal, dose-response and incidence concordance and no or few critical data gaps or conflicting data.	Moderate Confidence: Demonstrated dependent change in both events following exposure to a small number of specific stressors and some evidence inconsistent with an expected pattern that may be explained by factors such as experimental design, technical considerations, differences among laboratories, etc.	Low (Weak) Confidence: Limited or no studies reporting dependent change in both events following exposure to a specific stressor (i.e., endpoints never measured in the same study or not at all); and/or significant inconsistencies in empirical support across taxa and species that don't align with expected pattern for hypothesized AOP.

[1] Adapted from OECD AOP Handbook (OECD, 2014a,b,c), Annex 1. The BH considerations are rank ordered, as per Meek et al. (2014b) with respect for use in the overall weight of evidence determination of an AOP (i.e., biological plausibility > essentiality > empirical evidence). The following footnotes are verbatim from OECD AOP Handbook (OECD, 2014a,b,c), Annex 1, except as indicated with bracketed text.

[2] The guidance for "high", "moderate" and "low" draws on limited current experience. Additional delineation of the nature of relevant evidence in these broadly defined categories requires more experience with larger numbers of documented AOPs.

[3] "Direct evidence" implies specifically designed experiments to consider the relevant element. "Indirect evidence" normally relates to empirical support and is largely duplicative of Element 3 [empirical evidence].

[4] To the extent possible, each of the relevant Bradford Hill considerations is addressed for each of the KERs (biological plausibility and empirical support) and KEs (essentiality) and separate rationales provided.

[5] While the essentiality of each of the KEs is addressed separately, delineation of the degree of confidence is based on consideration of evidence for all of the KEs within the AOP and therefore, only one rationale is required.

[6] This is normally considered on the basis of tabular presentation of available data on temporal and dose-response aspects, in a template that documents the extent of support. See, for example, Meek and Klaunig (2010).

[7] Note that this relates to concordance of dose response, temporal and incidence relationships for KERs rather than the KEs; the defining question is not whether or not there is a dose response relationship for the KE but rather there is concordance with that for earlier and later KEs. This is normally demonstrated in studies with different types of stressors.

contributors of the OECD AOP Handbook (OECD, 2014a) considered the workshop outputs in revising the evaluation and documentation of levels of confidence by categorical ratings of high (H) or strong, moderate (M), and low (L) or weak, for KEs, KERs and the overall AOP. The evaluation of KERs necessitates an understanding of the scientific support for (1) the Key Event$_{upstream}$, (2) the Key Event$_{downstream}$ and (3) the relationship between these events. The summary table template found in Annex 1 of the OECD AOP Handbook (OECD, 2014a), which includes rank-ordered evolved BH considerations with defining questions and brief narrative guidance for assessing the relative level of confidence for biological plausibility, essentiality and empirical evidence was applied in the case examples presented here. Table 1 (adapted from OECD, 2014a) outlines, in a streamlined manner, the defining questions that were considered in developing the case examples for purposes of communicating the elements of WoE. AOP developers are directed to the OECD AOP Handbook (OECD, 2014a), which includes guidance, current templates and suggested evidence table formats for summarizing data in support of WoE evaluations.

The rank ordering is based on experience in MoA/species concordance analysis and includes biological plausibility and empirical support (dose-response, temporality, and incidence) for KERs and essentiality of KEs in the context of the AOP. In rank order, the evolved BH considerations are: (1) biological plausibility; (2) essentiality; and (3) empirical support (dose-response, temporality, consistency). While biological plausibility and empirical support is considered for each of the KERs, essentiality of the KEs is considered in the context of the overall AOP. The summary table template and associated text in the OECD AOP Handbook (OECD, 2014a) for each specific BH consideration are intended to promote consistency of evaluation within and across AOPs and to explicitly capture the rationale of AOP developers for the levels of confidence assigned to KEs and KERs. Although not depicted in Table 1, the summary table template in the OECD AOP Handbook (OECD, 2014a) includes entry space for summarizing the level of evidence, as appropriate for a specific AOP, for each KE and KER for each of the three BH considerations. The objective of the template is to clearly delineate the degree of confidence that can be determined based on thorough review of the available data in the context of the outlined considerations for low (weak), moderate and high (strong).

More transparent documentation of articulated considerations for WoE evaluation of AOPs should increase the collective understanding of the importance and value of such analyses to underpin the scientific confidence in specific regulatory applications. Through transparent consideration of specified elements based

on WoE guidance, it is anticipated that there will be significant benefit imparted from the clarity of the rationales provided for the KEs and KERs, including to contributors for refining AOPs from the broader community, peer reviewers, and AOP users, as well as those interested in learning from the experiences of others to develop new AOPs.

Based on previous experience in the development of WoE considerations, testing of proposed approaches is the most important element in their informed evolution. Thus, the objectives of the application of the case examples presented herein are to contribute to the further development of the OECD AOP Handbook (OECD, 2014a), to improve upon the current guidance by reducing ambiguities, eliminating unnecessary duplicative evaluations of supporting data and increasing transparency. The summaries presented here represent comparatively early experience in the development and documentation of applying the BH considerations to AOPs and these procedures are expected to continue to evolve, perhaps rapidly, as AOPs are developed in the near future.

3. Case examples

To illustrate application of WoE evaluation, analyses are provided for several AOPs:

- Aromatase inhibition leading to reproductive dysfunction in fish
- Arylhydrocarbon receptor (AhR) activation leading to induction of cytochrome P450 (CYP) monooxygenases and oxidative stress (see Annex)
- Juvenile hormone agonist-induction leading to increase in male offspring in the arthropod Cladocera (see Annex)
- Binding of certain organophosphates to neuropathy target esterase (NTE) leading to delayed neuropathy (see Annex)
- Agonist binding to estrogen receptor α (ERα) leading to an increased risk of endometrial cancer (see Annex)
- A chemical specific case example: induction of cytotoxicity and regenerative hyperplasia by oral hexavalent chromium (Cr(VI)) leading to duodenal tumors in mice – a chemical specific MoA analysis (see Annex)

The case example Aromatase Inhibition Leading to Reproductive Dysfunction in Fish, a relatively well-characterized AOP (https://aopkb.org/aopwiki/index.php/Aop:25), is presented in detail below. The other analyses are presented in the accompanying Annex. Evaluation of each KE and KER would be the norm for consideration of the WoE in the development of a complete AOP. Essentiality of KEs is also addressed in the context of all of the KEs in the AOP. However, to gain broader experience across a variety of AOPs, rather than focus on one AOP in detail, we elected to examine a subset of KEs and KERs in the six case examples.

3.1. Aromatase inhibition leading to reproductive dysfunction (in fish)

This AOP characterizes the consequences of inhibition of the enzyme cytochrome P450 aromatase (CYP19) – the MIE – relative to reproductive effects and, potentially, population-level responses in fish (Ankley et al., 2009a,b, 2010). The species used for much of the work deriving this AOP is the fathead minnow (*Pimephales promelas*), a model fish species in ecotoxicology research and regulatory applications (Ankley and Villeneuve, 2006). However, knowledge of basic comparative reproductive endocrinology, including cross-species conservation of CYP19 structure and function, suggests a relatively broad biological domain of applicability of this basic AOP not only to fishes but, potentially, other oviparous vertebrates as well (Celander et al., 2011;Norris and Carr, 2013). Briefly, aromatase catalyzes the conversion of testosterone (T) to 17β-estradiol (E2), which is involved in different aspects of reproduction in fish, including stimulation of production of vitellogenin (VTG; an egg yolk precursor protein) in the liver of females, through activation of the estrogen receptor (ER). Hepatic VTG enters the bloodstream, is taken up into the ovary, and incorporated into developing oocytes. A number of environmental contaminants, including some pesticides and drugs, can inhibit the activity of CYP19. Exposure of reproductively-active female fish to aromatase inhibitors decreases measured activity of ovarian CYP19, resulting in a cascade of downstream KEs, including a depression in plasma E2 concentrations, subsequent decreases in plasma VTG, lowered deposition of VTG into developing oocytes, and depressed egg production (fecundity) that can be translated, via modeling, into population declines (Fig. 2).

There are seven pairs of KEs (and associated KERs) that comprise the aromatase inhibition-reproductive dysfunction AOP (Fig. 2; Table 2). The first pair of KEs is inhibition of CYP19 activity (the MIE) resulting in decreased ovarian production of E2 (described by KER1); then, reduced plasma concentration of E2 (described by KER2); depressed VTG production in the liver (described by KER3); decreased plasma VTG concentrations (described by KER4); impaired oocyte development (described by KER5); reduced fecundity (described by KER6); and, finally, decreases in the population (described by KER7). Also indicated in Table 2 are WoE rankings for the KERs associated with the pairs of KEs in the AOP. Below, for illustrative purposes, we explore two quite different types of KERs from the AOP in terms of application of the WoE considerations summarized in Table 1 (which corresponds to Annex 1 of the OECD AOP Handbook (OECD, 2014a)). Specifically, we evaluate biological plausibility, and empirical evidence in assessing the relationship between decreases in the KEs of plasma E2 and VTG production (KER3), and between the KEs of fecundity and population status (KER7). As to essentiality of KEs, we discuss this below in terms of assessment of the overall AOP.

3.1.1. WoE Evaluation of KER3 (reduction in plasma 17β-estradiol concentrations leading to reduction in transcription and translation of vitellogenin)

3.1.1.1. Biological plausibility. Based on well-established knowledge of normal reproductive biology in oviparous animals, the biological plausibility of a relationship between plasma E2 concentrations and hepatic VTG production is high (strong). Vitellogenin synthesis in fish is localized in the liver and is well documented to be regulated by estrogens via interaction with ERs (Norris and Carr, 2013). There is extensive *in vitro* and *in vivo* evidence with multiple fish species of this relationship. This includes *in vivo* studies with females in which the status of the two parameters has been assessed relative to one another over the course of normal reproductive cycles (Norris and Carr, 2013), and mechanistic *in vitro* experiments showing that VTG production in hepatic tissues can be blocked by ER antagonists (e.g., Sun et al., 2010; Petersen and Tollefsen, 2012). Indirect evidence for biological plausibility of the relationship between E2 and VTG production comes from studies with male fish, which maintain the ability to produce VTG, but normally do not express the protein. Exposure of males to estrogens (including E2) results in induction of hepatic VTG production, a response not caused by exposure to other chemical classes or environmental stressors (Sumpter and Jobling, 1995).

3.1.1.2. Empirical support. Concordance of empirical observations of dose-response and temporal relationships between E2 and VTG depressions in fish also provides strong evidence for linkage of the two endpoints. For example, intensive time-course/dose-response studies with fathead minnows conducted with two

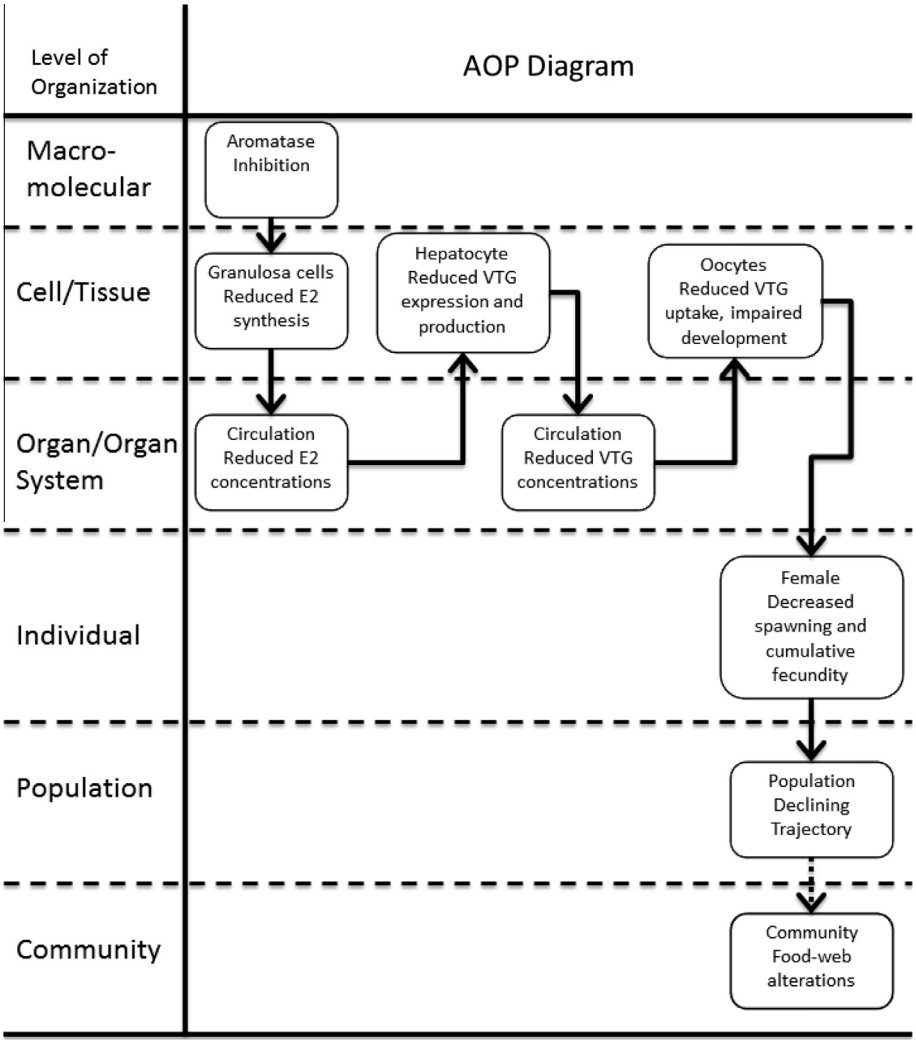

Fig. 2. Overview of an adverse outcome pathway relating inhibition of aromatase inhibition to reproductive dysfunction in fish. E2, 17β-estradiol; VTG, vitellogenin.

Table 2
KEs and WoE analysis of KERs for the adverse outcome pathway of aromatase inhibition leading to reproductive dysfunction in fish.

Key Event (upstream)	Key Event (downstream)	Weight-of-evidence for the KER
Aromatase inhibition	Ovarian (granulosa cell) E2 synthesis (reduction)	KER1: High (Strong)
Ovarian (granulosa cell) E2 synthesis (reduction)	Plasma 17β-estradiol concentrations (reduction)	KER2: High (Strong)
Plasma 17β-estradiol concentrations (reduction)	Transcription and translation of vitellogenin (reduction)	KER3: High (Strong)
Transcription and translation of vitellogenin (reduction)	Plasma vitellogenin concentrations (reduction)	KER4: High (Strong)
Plasma vitellogenin concentrations (reduction)	Vitellogenin uptake, impaired oocyte development (reduction)	KER5: Moderate
Vitellogenin uptake, impaired oocyte development (reduction)	Spawning and cumulative fecundity (reduction)	KER6: Moderate
Spawning and cumulative fecundity (reduction)	Population trajectory (decrease)	KER7: Moderate

known CYP19 inhibitors, fadrozole (a drug) and prochloraz (a conazole fungicide), showed that depressions in plasma E2 precede those of VTG, and both were dependent on chemical dose with the decrease in plasma E2 occurring at doses equal to or lower than those at which VTG decreased (Villeneuve et al., 2009; Ankley et al., 2009b). In terms of consistency of the relationship across multiple test systems, several other studies in fish have shown that chemicals which decrease plasma E2 also depress vitellogenesis. For example, Yu et al. (2014) recently reported that polybrominated diphenyl ethers depressed E2 production and hepatic VTG mRNA expression in zebrafish (*Danio rerio*). Although not included in this brief discussion, the OECD AOP Handbook (OECD, 2014a) suggests summarizing empirical evidence in tabular format to illustrate dose response, temporality and incidence concordance.

3.1.2. WoE of KER7 for the AOP of aromatase inhibition leading to reproductive dysfunction in fish

3.1.2.1. Biological plausibility. In terms of biological plausibility, it is intuitive that, in the absence of emigration, decreases in embryo production would depress population size. Consistent with this, Miller and Ankley (2004) describe a density-dependent, Leslie matrix model that utilizes empirical life-history data for the fathead minnow, where population trajectories are tightly coupled

to fecundity. Hence, in terms of plausibility KER7 would be scored high.

3.1.2.2. Empirical support. There is limited empirical evidence from the open literature that fathead minnow population size will decrease if fecundity is decreased by an endocrine-active chemical. Kidd et al. (2007) conducted a study in which an entire lake was treated with the ER agonist 17α-ethinylestradiol, and reported declines in fathead minnow population size corresponding with signs of reduced fecundity. However, while there is every expectation that decreased recruitment can decrease population size, many other variables also can affect population status (Kramer et al., 2011).

Overall, although the biological plausibility of the linkage between fecundity and population status for this AOP is strong (and there is some field evidence to support this), in the absence of additional empirical dose-response and time-course data, the WoE for KER7 is considered only to be moderate.

3.1.3. Evaluation of WoE for entire AOP (aromatase inhibition leading to reproductive dysfunction (in fish))

As noted in the OECD AOP Handbook (OECD, 2014a), WoE rankings for the individual KERs will heavily influence the overall WoE analysis for the entire AOP. However, there are components of the WoE analysis (e.g., essentiality) that are most amenable for assessment when considering the KEs (and associated relationships) in the context of the whole AOP. Further, some facets of other WoE considerations and associated guidance itself (OECD, 2014a) become apparent only when evaluating the entire AOP. For example, when evaluating the whole AOP, it is possible to assess relationships between KEs that are not immediately adjacent to one another, denoted as indirect KERs (OECD, 2014a), which can be important in the development of quantitative AOPs (in preparation, see Perkins et al. (in preparation), https://aopkb.org/saop/workshops/somma.html). Below we describe an analysis for the entire aromatase inhibition-reproductive dysfunction AOP, conducted using the considerations summarized by OECD (2014a). It should be noted that, due to space limitations, this evaluation is meant to be illustrative rather than exhaustive, so only a comparatively small subset of available literature/data supporting different aspects of the AOP are utilized.

Biological plausibility of the AOP is, as a whole, considered strong. Based on knowledge derived from literally hundreds of studies concerning normal fish reproductive endocrinology and population dynamics published in the open literature (and summarized in text books such as Norris and Carr (2013)), the cascade of events depicted in Fig. 2 is both plausible and well documented. Further, based on comparative endocrinology research, it is highly probable that this AOP has a broad domain of applicability in terms of its relevance, at least to fish species, if not other oviparous vertebrates (Celander et al., 2011; Norris and Carr, 2013).

Overall, based on studies with multiple chemicals, sampling times, and fish species, empirical support of this AOP is judged to be strong. There are extensive empirical data in support of the overall AOP. Several studies with known CYP19 inhibitors (both drugs and pesticides) have demonstrated a dose-dependent relationship between one or more of the KEs in the AOP and impacts on fecundity. For example:

(1) In 21-day fathead minnow reproduction studies with fadrozole, prochloraz, and propiconazole, dose-dependent decreases in plasma E2 were associated with corresponding depressions in plasma VTG in females, and subsequent decreases in fecundity (Ankley et al., 2002, 2005; Skolness et al., 2013).
(2) Both dose-dependency and consistency of the AOP in an additional fish species were demonstrated in work with the Japanese medaka (*Oryzias latipes*) in 21-day studies with letrozole, a pharmaceutical specifically designed to inhibit CYP19 (Sun et al., 2007). In those studies, decreases in plasma VTG concentrations in females corresponded, in a dose-dependent manner, with depressed egg production in the fish and the decreases in VTG occurred at doses equal to or greater than doses which depressed egg production.
(3) There also are empirical linked temporal-dose data supporting the aromatase inhibition-reproductive dysfunction AOP in fish. For example, Villeneuve et al. (2009) and Ankley et al. (2009b) described experiments with fadrozole and prochloraz, respectively, in which fathead minnows were exposed to two concentrations of the CYP19 inhibitors and sampled at multiple times after exposure was initiated. Although only a subset of the KEs was measured, results observed were nonetheless consistent with the proposed AOP.

Specifically, in exposed females, depressions in ovarian production of E2 preceded decreased plasma levels of E2, which preceded decreased plasma VTG concentrations. Further, this set of relationships reflected the concentration of the chemical (fadrozole or prochloraz) to which the fish were exposed, with responses occurring more quickly and to a relatively larger degree in the high- versus low-dose treatment groups (Villeneuve et al., 2009; Ankley et al., 2009b).

A potentially critical component of the WoE analysis for an overall AOP involves evaluation of the concept of essentiality, i.e., demonstration that if a KE (which could include the MIE) in an AOP is blocked in some manner, downstream KEs including the AO do not occur. This type of evidence can be generated using chemical inhibitors of upstream KEs or, in some cases, knockout animal models that have been genetically-modified such that a functional component of the pathway of interest is lacking. Essentiality also can be demonstrated by reversibility of an impact, i.e., when the chemical stressor impacting an upstream event is removed, subsequent KEs recover.

Evidence of essentiality for the aromatase inhibition-reproductive dysfunction AOP arises both from recovery studies, and from a somewhat novel example of reversibility related to biological compensation/adaptation. The multiple time-point/dose study design described above (Villeneuve et al., 2009; Ankley et al., 2009b; Ankley and Villeneuve, 2015) included a recovery phase where the fathead minnows were sampled for a period of time after the chemical exposures were stopped. In those studies, measurements of KEs in fish subsequently held in clean water after the chemical exposures ended exhibited temporal relationships consistent with the proposed AOP, i.e., depressed ovarian synthesis of E2 recovered (increased) first, followed by plasma E2 concentrations, and then plasma VTG levels. Somewhat unexpectedly, however, there was additional evidence for essentiality of these AOP KEs during the actual chemical exposure. Specifically fish, most often in the low dose group, after an initial depression in synthesis and plasma concentrations of E2, exhibited compensatory-type responses that also reflected the anticipated temporal relationships reflected in the AOP, e.g., increases in ovarian synthesis of E2 preceded recovery of plasma E2 concentrations. This compensatory response appeared to be due to the up-regulation of key steriodogenic enzymes (including CYP19) in the fish due to negative feedback within the hypothalamic-pituitary-gonadal axis. Notably, similar time-course/dose-response studies with other sex steroid synthesis inhibitors in the fathead minnow have produced compensation and recovery patterns analogous to the fadrozole and prochloraz work (e.g., Ankley and Villeneuve, 2015).

Taken as a whole, these studies demonstrating relationships between early KEs in the AOP in the context both of recovery and compensation provide powerful support of essentiality. What is lacking in the literature is extension of analysis of compensation/recovery to later KEs in the AOP, most notably the AO of fecundity decreases.

Overall, based on WoE rankings of the individual KERs (Table 2), and the WoE for the entire pathway, the aromatase-inhibition reproductive dysfunction AOP (Fig. 2) is considered strong, although uncertainties remain in terms of prediction of population-level responses. This arises more from a lack of data, rather than contradictory evidence of a disconnection between fecundity and population size. Consequently, from a practical perspective in terms of the WoE relative to uncertainty, this AOP would be judged as strong for the AO of fecundity decreases and moderate for the AO of population-level effects. However, it could be argued that this uncertainty is more quantitative than qualitative, as it is obvious that if there is no reproduction, in the absence of emigration, population extinction will occur.

4. Lessons learned in applying the evolved WoE framework

During the discussions at the workshop, considerable focus was given to clarifying the descriptors to guide decision making for assigning confidence in components of WoE for KEs, KERs and the overall AOP. The workgroup also contributed to clarification of terminology in development of the OECD Users' Handbook. The "lessons learned" based on the experience of the workgroup in discussions and during the development and analysis of the diverse set of the case examples (see Annex) are articulated below, both in recognition that evolution is informed principally by practical application and to aid others who are embarking on documentation of WoE evaluation of AOPs. For a more complete picture of the evolution of AOP development and current thinking on application to integrated approaches to testing and assessment, readers should consult the companion publications stemming from the Somma Lombardo workshop: Garcia-Reyero (2015), Groh et al. (2015a,b), Tollefsen et al. (2014), Villeneuve et al. (2014a,b), and Perkins et al. (in preparation) (see https://aopkb.org/saop/workshops/somma.html).

4.1. The need to focus on KERs

The evolved WoE framework illustrated in Table 1 (and expanded in the OECD AOP Handbook (OECD, 2014a)) clarifies the need for considering both KEs and KERs in addressing WoE for AOPs, with essentiality relating to KEs (within the context of the AOP) and biological plausibility and empirical support relating to KERs. As the framework evolved and discussions proceeded, the emphasis relevant to consideration of WoE in relation to intended uses of AOPs increased focus on understanding and defining $KE_{upstream}$ so as to enable prediction of $KE_{downstream}$ using results of assays that measure $KE_{upstream}$ and suitable prediction models. For example, one may wish to use results of the KEs measured with *in vitro* assays of receptor binding and transcriptional activation in an estrogen pathway to predict *in vivo* responses of the uterotrophic assay (Rotroff et al., 2013), an *in vivo* Tier 1 screening assay in the USEPA Endocrine Disruptor Screening Program. To do so, requires a predictive relationship that uses results of the KEs of binding and transactivation as the input to a model, whose output is the estimated response in the uterotrophic assay. While the overall AOP represents the sequence of KEs, it is the KERs which characterize the relationships between KEs, and in essence, such relationships are qualitative or quantitative prediction models. Hence, in the current framework, distinction of various components of KERs, which encompass knowledge of $KE_{upstream}$ and $KE_{downstream}$, the understanding of the relationship (the ability to predict $KE_{downstream}$ from $KE_{upstream}$) between KEs has been made more explicit. Capturing the evidence for causality at the level of each KER provides two additional advantages. First, it promotes a systematic approach to the evaluation and recording of the evidence by directing the author to consider each link in the chain before evaluating the overall AOP. Second, it facilitates the reuse of the evidence in other AOPs for which the specific KER is a component thereby promoting consistency and completeness in describing all AOPs.

4.2. Biological plausibility of KEs and KERs

For evaluating biological plausibility, emphasis is placed on the biological basis of the AOP and its KEs and KERs. The initial impulse of most scientists to questions of biological plausibility of the KEs and KERs within an AOP is to look to experimental evidence to make a cogent case. In particular, there is a predilection for toxicologists to review and cite experimental evidence of dose response, effects and outcomes of chemical exposure when undertaking the WoE evaluation for all three elements of the AOP WoE framework (biological plausibility, essentiality and empirical evidence). This initial experience within the workshop contributed to evolution of the defining questions and the descriptors to guide decision making for assigning level of evidence. The interactions between experts in human health and experts in environmental and ecotoxicology were particularly valuable.

In addressing the defining question in the OECD AOP Handbook summary table template (e.g., Table 1) to determine the WoE of biological plausibility of an AOP, focus needs to be placed on evaluating the extent to which the AOP aligns with current understanding of normal physiology, biological processes and pathways of pathogenesis culminating with adverse effects. In other words, characterizing the extent to which the scientific community has accepted an established mechanistic basis for the steps in the pathway. The AOP developer should explicitly address the understanding of the biology, i.e., the degree of knowledge and extent of scientific acceptance of the biological steps and sequence in the AOP. Although the foundation in part for such acceptance will be empirical evidence, the rationale for plausibility should focus on biological pathways and relationships of KEs; detailed analysis of dose response results is conducted in the empirical evidence evaluation step. Since the AOP is describing the biological basis for the adverse outcome (AO), information from non-chemical perturbations can also be used in establishing the biological plausibility. For example, knockout mouse models that block a specific KE can establish both the causal relationships necessary for biological plausibility as well as the essentiality of the KE as described below.

Increasingly, biological plausibility is being limited to extent of understanding/acceptance of the biology to distinguish it from essentiality of key events. While the overlap is recognized, to some degree, this is a function more of "codifying" WoE considerations to simplify and increase common understanding.

4.3. Essentiality of the KEs within the AOP

As was the case for biological plausibility, the initial discussion of the workgroup at the workshop tended to focus on reviewing and citing experimental evidence of dose response and temporal concordance (i.e., empirical support) to evaluate essentiality, e.g., making the case that $KE_{downstream}$ is essential, as indicated by the fact that there is a dose dependent change in $KE_{downstream}$ as a result of the dose response of $KE_{upstream}$. It became clear that to better differentiate essentiality from empirical evidence, and to more

clearly address this element of the WoE framework, there was a need to more explicitly clarify the nature of supporting data and descriptors guiding determination of the level of evidence associated with high, moderate and low confidence.

For essentiality, as reported in Meek et al. (2014b), the most persuasive evidence is when the $KE_{downstream}$ is not observed when $KE_{upstream}$ is blocked. Going into the workshop, studies that demonstrate reversibility (e.g., stop or recovery experiments,) were identified as the sort of evidence that provides a high degree of certainty regarding essentiality. During the workshop, participants pointed out that in pharmacology, the use of an antagonist to block the response elicited by an agonist is a classical method for establishing essentiality. In addition to classical pharmacological antagonism, evidence from genetically engineered models (e.g., knockout, knock-in, conditional gene modifications, etc.) can provide a high degree of specificity and certainty regarding essentiality. The OECD AOP Handbook workgroup determined that indirect evidence, in particular, modulation of a KE by a factor that attenuates or augments the response is also useful, and should be viewed as providing a moderate degree of certainty for essentiality. In the developing discussion of the components of WoE, it was also clarified that since essentiality of the KEs relates to prevention or attenuation of any downstream key event when an upstream key event is blocked or modified, it is most readily assessed in the context of the entire AOP, rather than individual KERs.

In an ideal case, data would be available to allow evaluation of the essentiality of each and every KE. However, this would require an enormous amount of experimental evidence. From a practical perspective, experience has shown that essentiality data typically focus on one or a limited set of KEs. Thus, in evaluating essentiality, the determination is made for the overall AOP. For example, to demonstrate the essentiality of the KE of translation to the overall AOP, cycloheximide could be used to block protein synthesis. For AOPs focused on tumorigenesis, stop exposure studies in cancer bioassays and initiation promotion studies can also provide evidence of essentiality of a KE to the overall AOP.

4.4. Empirical support

The case examples emphasized the importance of separating empirical support from biological plausibility and its lesser weighting in contributing to WoE; experience gained was particularly critical in this context since empirical support is normally demonstrated in studies with different types of stressors. Application of the guidance to the case examples was also helpful in increasing understanding that empirical support relates to "concordance" of dose response, temporal, and incidence relationships for KERs rather than the KEs. Although the initial focus was on whether or not there was a dose-response relationship for a defined key event, the emphasis was shifted to reviewing and citing experimental evidence of agreement in dose response and temporality among key events (i.e., concordance), which is demonstrated based on studies with administered stressors. For dose response, this would be determination that the KER reflects dose-dependent changes in $KE_{downstream}$ as a result of the dose response of $KE_{upstream}$. For temporality, demonstration that $KE_{upstream}$ precedes $KE_{downstream}$ was expanded to include evidence that $KE_{upstream}$ occurs at lower doses, at earlier time points and at higher incidence than $KE_{downstream}$. For concordance, there was agreement that the strongest degree of certainty arises from multiple studies in relevant test systems (i.e., orthogonal methods) which show similar qualitative/quantitative responses. The OECD AOP Handbook contributors further refined the questions and evidence descriptors, and importantly, provided greater granularity to the descriptors for M and L evidence descriptors.

The contribution of the tabulation of temporal, dose-response and incidence data, in a format to identify relevant trends identified in the Guidance was also recognized and recommended in the OECD AOP Handbook (OECD, 2014a). Although beyond the scope of the purpose and intent of the case example summaries presented here, it is acknowledged that standardized presentation in tabular format has the potential to simplify considerably the review of supporting evidence to support WoE determinations for AOPs.

4.5. Inconsistencies and uncertainties

One of the key lessons learned for those developing or evaluating AOPs is that uncertainties and inconsistencies need to be explicitly considered and integrated as part of the determination for each element of the three recommended BH considerations: biological plausibility, essentiality and empirical evidence. Originally, WoE guidance from OECD included a separate element to address uncertainties and inconsistencies, and this element was initially carried through in the evaluation of the case examples presented here. However, even at that time it was recognized that this was largely redundant, in that inconsistencies and uncertainties were implicitly already factored into the evaluations of each of the other elements of the WoE framework as part of the process for determining the degree of evidence (e.g., H, M or L). Following the workshop, in the process of modifying the OECD AOP Handbook, this lesson learned contributed to deletion of a separate element for inconsistencies and uncertainties to avoid redundancy.

4.6. The role of chemical-specific case example AOPs

There is a dynamic tension in developing AOPs between the need to rely on, and use, experimental evidence and knowledge derived from specific chemicals and the need to construct an AOP such that it focuses on the biological steps and pathways, in a manner that is agnostic to a specific chemical (Villeneuve et al., 2014a). For well characterized pathways and MoAs, it may be easier to refrain from over reliance on a specific chemical. But, recognizing that this will not always be the case in terms of available data, and in order to foster AOP development, including development of more speculative AOPs, it is acceptable within the OECD AOP program to propose either of two types of AOPs: (1) those with an MIE that could be affected by a variety of chemicals, and (2) those applicable to a single chemical or a very limited number of chemicals (OECD, 2014a). For such chemical-specific case examples, the WoE evaluation will typically be limited to a specific MoA analysis of a single chemical. With time and with additional input (e.g., additional data and analyses encompassing a broader domain of chemistries) the breadth of application of these chemical-specific case example AOPs could expand.

4.7. Consideration of non-adjacent KERs

As discussed in the OECD AOP Handbook (OECD, 2014a), an indirect KER is one that links a pair of non-sequential KEs. For example, if there is uncertainty or difficulty in measuring KE_{n+1}, one could use an indirect KER that links the upstream KE_n to the downstream KE_{n+2}. In other words there may be sufficient data to establish the KE_{n+1} as part of the AOP, but to evaluate the WoE of the AOP one may have to rely on the indirect KER by "leaping over" KE_{n+1}. When evaluating indirect KERs, the WoE for these relationships should be explicitly described and knowledge gaps communicated. This will help those interested in applying the AOP for a specific use to ascertain whether measures of responses at KE_n can be used to predict downstream KEs (including the AO),

with sufficient confidence to support such a use, or whether the uncertainty is too great to support bypassing KE_{n+1}. Such decisions will clearly depend upon the intended use of the AOP; for example, we expect that certain uses, such as priority setting, may have a tolerance for greater uncertainty than other regulatory applications.

4.8. Challenges in presenting and communicating WoE rationales

The analyses conducted to illustrate application of the WoE approach for the case example AOPs highlight the difficulties encountered when trying to condense and summarize considerable quantities of scientific evidence, often of varying quality and developed for diverse purposes, into succinct lines of evidence which align with the three BH considerations of biological plausibility, essentiality and empirical evidence. One element that was not discussed here, but that is addressed elsewhere and is relevant to establishing scientific confidence in an AOP, is the issue of test method validity with respect to data quality and reliability. Although regulatory testing under the auspices of OECD or EPA test guidelines employs validated methods, much of the data that will be used in constructing AOPs will be based on non-guideline methods, and OECD has developed draft guidance (OECD, 2014c) to aid developers and reviewers to evaluate the quality of data produced by such methods. Cox et al. (2014) also discuss an approach to, and the importance of, evaluating the analytical performance of assays used to measure KEs and prediction models (e.g., qualitative or quantitative KERs) and communicating the scientific rationale explaining the confidence in use of the KER for a fit for purpose application. Moving forward, use of the summary table template in Annex 1 of the OECD AOP Handbook (OECD, 2014a) should help to improve presentation and communication of the evidence underpinning the H, M, L designations for the WoE for each of the three recommended BH considerations. As more experience is gained it is expected that this too will contribute to continued refinement overall. The recent development of the AOP "*alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations*" by Yauk et al. (2015), which was also used in an iterative process to further test and refine the OECD AOP Handbook in its entirety, provides an example of an in-depth analysis of a data rich AOP. This AOP clearly illustrates enhanced transparency that is required in presenting supporting data and evaluation of WoE based on lessons learned and the continuing efforts under the OECD AOP Programme and supporting initiatives.

5. Looking ahead: quantitative WoE evaluation – challenges and opportunities

Weight-of-evidence evaluation involves a diverse set of methods. Transparency and reproducibility of the results are increasingly important areas of concern within the scientific community, though achieving these objectives is challenging. Although the OECD AOP Handbook (OECD, 2014a) provides guidance for improved transparency and enhances communication of the rationales for arriving at H, M or L determinations in WoE evaluations of AOPs, reproducibility remains a challenge. The qualitative WoE logic method recommended in the OECD AOP Handbook may be useful for simple AOPs. However, qualitative methods are limited by the inability to easily deal with complex datasets where multiple criteria and metrics make it difficult to develop a narrative that integrates multiple logical constructs together. Consequently, a quantitative approach for AOP WoE evaluation using multi-criteria decision analysis (MCDA) (Linkov et al., 2009; Perkins et al., in preparation) holds considerable promise.

Quantitative WoE methods require developing criteria (through which experts judge the importance of individual lines of evidence) and metrics (used to measure the performance on criteria), through which alternatives are evaluated (Linkov et al., 2009, Linkov and Moberg, 2012). In MCDA, criteria, metric weights, and scores are used to represent either objective data or the subjective preference values of decision makers with respect to decision goals. Weights are assigned to criteria and metrics based on their relative importance for the ultimate evaluation goals and individual lines of evidence are subsequently scored based on their relative performance on these criteria and metrics. The weights and scores are then synthesized using different mathematical or statistical models to form an overall conclusion. Quantitative WoE methods can be based on individual judgment or group decision process.

In the context of AOPs, advanced quantitative methods such as MCDA and Bayesian Network analysis are especially useful. The evolution of the BH considerations for WoE for MoA and AOPs (Table 1) reflects the order of relative importance of each consideration (e.g., biological plausibility > essentiality > empirical evidence) and for each consideration, the degree of confidence in KERs (i.e., H, M, or L). This hierarchical approach lends itself to decision modeling and quantitative scoring, wherein the overall WoE can be calculated as integrative metrics of weights associated with the BH considerations and their associated KER confidence levels.

For illustration purposes, we developed a prototype quantitative MCDA model (Fig. 3) for the AOP of aromatase inhibition leading to reproductive dysfunction in fish (described above). The MCDA model description generated for the quantitative WoE evaluation of this AOP are presented in the Supplementary Material. Input was provided by expert elicitation from two leading experts involved in the development of this AOP. DECERNS software (Linkov and Moberg, 2012) was used for the analysis, and the supplementary figures presented are direct outputs of the DECERNS Software.

It is important to note that the purpose of the following discussion is to be illustrative only and should be taken as a prototypical demonstration of MCDA applied to WoE for an AOP; the quantitative scores are not to be used in any manner that would suggest they are definitive. The quantitative MCDA model was constructed using the three evolved BH considerations (biological plausibility, essentiality and empirical evidence), their relative importance (biological plausibility > essentiality > empirical evidence), and the scores of H, M and L, evaluated through defining questions evaluated through the defining questions (Table 1). The model allows for evaluation of relative strength of evidence underpinning KEs and KERs within this AOP and follows the logic set forth in the OECD AOP Handbook and the WoE summary table template (OECD, 2014a).

The ranking of KERs is shown in Fig. 4. This prototype MCDA model shows that the highest WoE confidence is in KER1 (aromatase inhibition leading to reduced E2 synthesis) and KER2, (reduced E2 synthesis leading to reduced plasma E2). The lowest WoE confidence is found to be in KER6 (reduced oocyte growth leading to reduced fecundity).

A quantitative WoE approach such as MCDA allows for incorporation of inputs from multiple experts, strengthens WoE logic constructs by adding the visual effect of a mapped decision structure, includes quantitative weighing of individual lines of evidence, and allows for sensitivity analysis. Exploration of the sensitivity of the WoE for each KER to changes in weighting schemes is illustrated in the Supplementary Material. Such a sensitivity analysis can be used in a value of information analysis and play an important role in delineating additional research areas where developing a greater

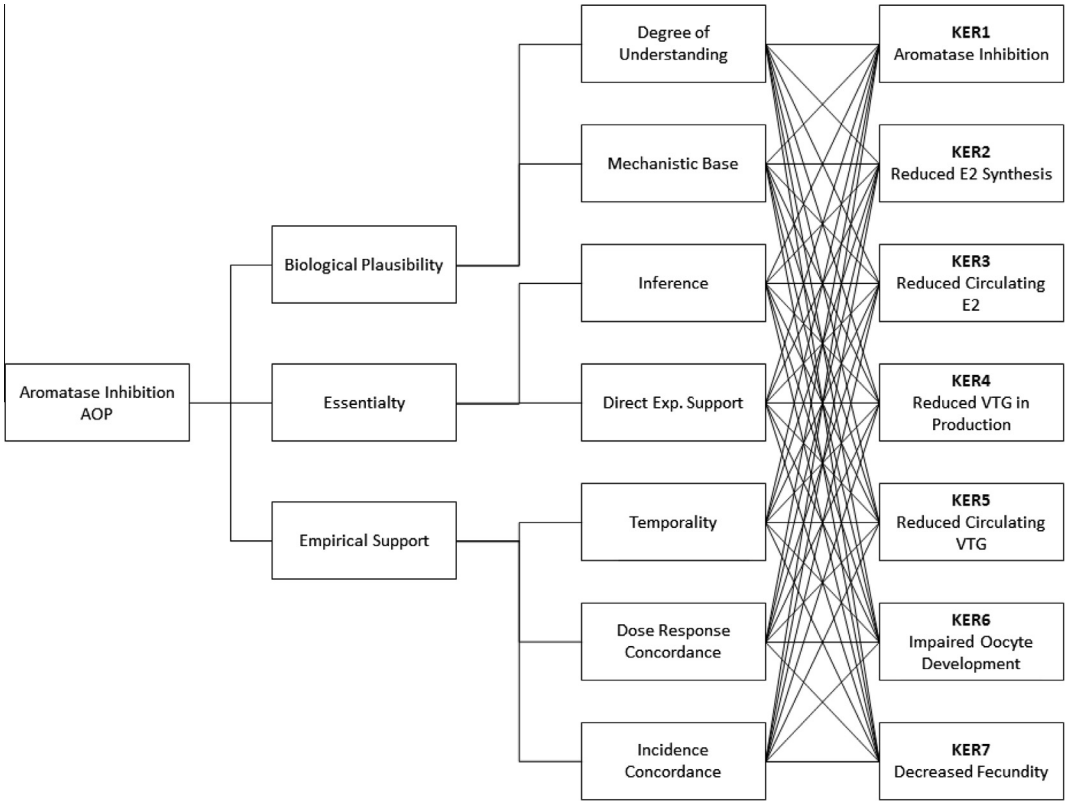

Fig. 3. Prototype MCDA model for quantitative WoE determination for KERs in the aromatase AOP.

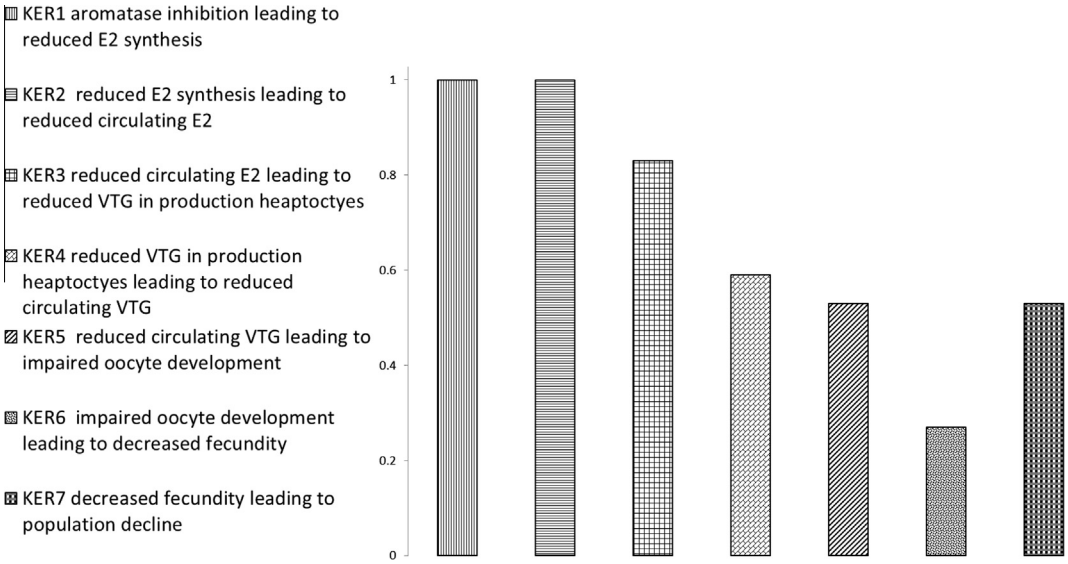

Fig. 4. Scores for confidence assessment in each KER of the aromatase inhibition AOP. Each KER contains an overall score from 0 to 1 that is a cumulative score originating from individual scores for all criteria.

degree of confidence in a KE or KER would have the greatest impact on improving the overall confidence in the AOP.

6. Conclusions

There are a large number of possibilities for using the WoE confidence levels for KEs and KERs and AOPs. At the workshop, and in subsequent follow up discussions, approaches to combining the hierarchical weighting of the three BH considerations with the individual confidence weightings for each KER to arrive at an overall conclusion about the confidence level of the overall AOP were considered. As illustrated, MCDA can be used to construct a quantitative model for accomplishing this. However, at this early stage of WoE evaluations of AOPs, we recommend that AOP developers conduct a point-by-point qualitative evaluation of the overall

AOP that is supported by the summary table, and also consider including a written narrative rationale in terms of overall level of confidence (i.e., H, M or L) in the AOP. We also recommend that this rationale be uploaded into the AOP-Wiki, to ensure transparency and promote collaborative discussions across both the regulatory and scientific community. Once experts have gained adequate experience in applying evolved BH considerations to AOPs, and likely not in the too distant future, it will be feasible to organize a consensus expert elicitation process to refine quantitative MCDA application to AOP WoE determinations.

It is important to note that there will likely be cases where the overall confidence in using a KE or KER to predict the AO would be low. Nevertheless the confidence in specific KEs or KERs may be sufficient to use for a particular assessment application. For example, the MIE or an early KER may be known with sufficient certainty that assays for these could be used for purposes of priority setting or read across even when the knowledge of the quantitative predictivity of these KEs for the AO may be too uncertain for use in hazard identification or risk assessment purposes. In the AOP-Wiki, the WoE documentation provides degree of confidence in KERs that can assist with such decisions; the network view illustrates the degree of confidence graphically by the weight of the arrows linking KEs.

The quantitative MCDA prototype illustrated here holds considerable promise to enhance the rigor, transparency and reproducibility for AOP WoE determinations. Compared to a qualitative approach, a MCDA approach enables more precise, transparent and explicit delineation of expert judgment. Even though defining criteria and metrics for specific AOPs can be relatively easy, reliable scoring and weighting may be quite challenging. A meaningful correlation of expert-driven weighting and scoring will require significant efforts from the expert community (Linkov et al., 2015).

Transparency Document

The Transparency document associated with this article can be found in the online version.

Acknowledgments

The content of this paper is the result of an international expert workshop on *Advancing Adverse Outcome Pathways (AOPs) for Integrated Toxicology and Regulatory Applications* held March 2-7, 2014 in Somma Lombardo, Italy. Support for the workshop was provided by the American Chemistry Council, BioDetection Systems, European Centre for Ecotoxicology and Toxicology of Chemicals, Environment Canada, European Commission Directorate General Joint Research Center, Human Toxicology Project Consortium, International Life Sciences Institute - Health and Environmental Science Institute, The Research Council of Norway (Grant no. 221455), Society for Environmental Toxicology and Chemistry, US Army Engineer Research and Development Center, and the US Environmental Protection Agency. The authors acknowledge the workshop organizing committee (Rick Becker, Natàlia Garcia-Reyero, Ksenia Groh, Marlies Halder, Sean Kennedy, Teresa Lettieri, Edward Perkins, Knut Erik Tollefsen, Bart van der Burg, Dan Villeneuve, and Maurice Whelan), and all workshop participants for both plenary and informal discussions that influenced the content presented here. The views expressed are those of the authors, and do not necessarily represent the views of the organizations the authors are affiliated with or the sponsors. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.yrtph.2015.04.004.

References

Ankley, G.T., Kahl, M.D., Jensen, K.M., Hornung, M.W., Korte, J.J., Makynen, E.A., Leino, R.L., 2002. Evaluation of the aromatase inhibitor fadrozole in a short-term reproduction assay with the fathead minnow (Pimephales promelas). Toxicol. Sci. 67, 121–130.

Ankley, G.T., Jensen, K.M., Durhan, E.J., Makynen, E.A., Butterworth, B.C., Kahl, M.D., Villeneuve, D.L., Linnum, A., Gray, L.E., Cardon, M., Wilson, V.S., 2005. Effects of two fungicides with multiple modes of action on reproductive endocrine function in the fathead minnow (Pimephales promelas). Toxicol. Sci. 86 (2), 300–308.

Ankley, G.T., Villeneuve, D.L., 2006. The fathead minnow in aquatic toxicology: past, present and future. Aquat. Toxicol. 78, 91–102.

Ankley, G.T., Bencic, D., Breen, M., Collette, T.W., Conolly, R., Denslow, N.D., Edwards, S., Ekman, D.R., Jensen, K.M., Lazorchak, J., Martinovic, D., Miller, D.H., Perkins, E.J., Orlando, E.F., Garcia-Reyero, N., Villeneuve, D.L., Wang, R.-L., Watanabe, K., 2009a. Endocrine disrupting chemicals in fish: developing exposure indicators and predictive models of effects based on mechanisms of action. Aquat. Toxicol. 92, 168–178.

Ankley, G.T., Bencic, D.C., Cavallin, J.E., Jensen, K.M., Kahl, M.D., Makynen, E.A., Martinovic, D., Mueller, N.D., Wehmas, L.C., Villeneuve, D.L., 2009b. Dynamic nature of alterations in the endocrine system of fathead minnows exposed to the fungicide prochloraz. Toxicol. Sci. 112, 344–353.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29, 730–741.

Ankley, G.T., Villeneuve, D.L., 2015. Temporal changes in biological responses and uncertainty in assessing risks of endocrine-disrupting chemicals: Insights from intensive time-course studies with fish. Toxicol. Sci., In Press

CCA, 2012. Integrating emerging technologies into chemical safety assessment: the expert panel on the integrated testing of pesticides. Council of Canadian Academies.

Celander, M.C., Goldstone, J.V., Denslow, N.D., Iguchi, T., Kille, P., Meyerhoff, R.D., Smith, B.A., Hutchinson, T.H., Wheeler, J.R., 2011. Species extrapolation for the 21st century. Environ. Toxicol. Chem. 30, 52–63.

Dellarco, V., Henry, T., Sayre, P., Seed, J., Bradbury, S., 2010. Meeting the common needs of a more effective and efficient testing and assessment paradigm for chemical risk management. J. Toxicol. Environ. Health B Crit. Rev. 13 (2–4), 347–360.

Council of Labor Affairs, Taiwan. 2012. Existing Chemical Substance Inventory. Chemical Substance Nomination & Notification, Council of Labor Affairs, Executive Yuan, Taiwan. Available at: <http://csnn.cla.gov.tw/content/englishHome.aspx>.

Cox, L.A., Popken, D., Marty, M.S., Rowlands, J.C., Patlewicz, G., Goyak, K.O., Becker, R.A., 2014. Developing scientific confidence in HTS-derived prediction models: lessons learned from an endocrine case study. Regul Toxicol Pharmacol. 69, 443–450.

European Commission. 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC,93/67/EEC, 93/105/EC and 2000/21/EC. Off. J. Eur. Union L396, 1–849.

Garcia-Reyero, N., 2015. Are adverse outcome pathways here to stay? Environ. Sci. Technol. 49 (1), 3–9.

Groh, K.J., Carvalho, R.N., Chipman, J.K., Denslow, N.D., Halder, M., Murphy, C.A., Roelofs, D., Rolaki, A., Schirmer, K., Watanabe, K.H., 2015a. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. Chemosphere 120, 764–777. http://dx.doi.org/10.1016/j.chemosphere.2014.09.068.

Groh, K.J., Carvalho, R.N., Chipman, J.K., Denslow, N.D., Halder, M., Murphy, C.A., Roelofs, D., Rolaki, A., Schirmer, K., Watanabe, K.H., 2015b. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: II. A focus on growth impairment in fish. Chemosphere 120, 778–792. http://dx.doi.org/10.1016/j.chemosphere.2014.10.006.

Hill, A.B., 1965. The environmental and diseases: Association or causation? Proceed. Roy. Soc. Med., 58: 295–300

Hughes, K., Paterson, J., Meek, M.E., 2009. Tools for the prioritization of substances on the domestic substances list in Canada on the basis of hazard. Regul. Toxicol. Pharmacol. 55, 382–393.

Kidd, K.A., Blanchfield, P.J., Mills, K.H., Palace, V.P., Evans, R.E., Lazorchak, J.M., Flick, R.W., 2007. Collapse of a fish population after exposure to a synthetic estrogen. Proc. Natl. Acad. Sci. 104, 8897–8901.

Kramer, V.J., Etterson, M.A., Hecker, M., Murphy, C.A., Roesijadi, G., Spade, D.J., Spromberg, J.A., Wang, M., Ankley, G.T., 2011. Adverse outcome pathways and ecological risk assessment: bridging to population-level effects. Environ. Toxicol. Chem. 30, 64–76.
Linkov, I., Loney, D., Cormier, S., Satterstrom, K.S., Bridges, T., 2009. Weight-of-evidence evaluation in environmental management: review of qualitative and quantitative approaches. Sci. Total Environ. 407, 5201–5207.
Linkov, I., Moberg, E., 2012. Multi-Criteria Decision Analysis: Environmental Applications and Case Studies. CRC Press, ISBN 978143985318.
Linkov, I., Massey, O., Keisler, J., Rusyn, I., Hartung, T., 2015. From weight of evidence analysis to quantitative data integration tools: the path to improve both confidence and utility of hazard assessments. ALTEX: Altern. Anim. Exp. 32, 5–10.
Lowell Center for Sustainable Production, 2012. Chemicals Policy: Other Nations [Internet]. Chemicals Policy & Science Initiative, Lowell Center for Sustainable Production, University of Massachusetts Lowell, Lowell, Massachusetts. Available at: <http://www.chemicalspolicy.org/chemicalspolicy.othernations.php>.
Meek, M.E., Armstrong, V.C., 2007. The assessment and management of industrial chemicals in Canada. In: Van Leeuwen, K., Vermeire, T. (Eds.), Risk Assessment of Chemicals: An Introduction, 2nd ed. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 591–621.
Meek, M.E., Klaunig, J.E., 2010. Proposed mode of action of benzene-induced leukemia: interpreting available data and identifying critical data gaps for risk assessment. Chem. Biol. Interact. 184 (1-2), 279–285.
Meek, M.E., Boobis, A.R., Cote, I., Dellarco, V., Fotakis, G., Munn, S., Seed, J., Vickers, C., 2014a. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. 34, 1–18.
Meek, M.E., Palermo, C.M., Bachman, A.N., North, C.M., Lewis, R.J., 2014b. Mode of action human relevance (MOA/HR) framework – evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J. Appl. Toxicol. 34, 595–606.
Miller, D.H., Ankley, G.T., 2004. Modeling impacts on populations: fathead minnow (Pimephales promelas) exposure to the endocrine disruptor 17β-trenbolone as a case study. Ecotoxicol. Environ. Safety 59, 1–9.
Mitchell, J., Pabon, N., Collier, Z.A., Egeghy, P.P., Cohen-Hubal, E., Linkov, I., Vallero, D.A., 2013. A decision analytic approach to exposure-based chemical prioritization. PLoS ONE 8, e70911.
Norris, D.O., Carr, J.A., 2013. In: Vertebrate Endocrinology, 5th ed. Elsevier, p. 580.
NRC, 2007. Toxicity testing in the 21st century: a vision and a strategy. National Research Council. Washington DC. Retrieved from <http://dels.nas.edu/resources/static-assets/materials-based-on-reports/reports-in-brief/Toxicity_Testing_final.pdf>.
NRC, 2010. Toxicity pathway-based risk assessment: preparing for paradigm change, a symposium summary. National Research Council.
OECD, 2013. Guidance Document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment No. 184. Available at: <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en>.
OECD, 2014a. Users' handbook supplement to the guidance document for developing and assessing AOPs. Available at: <https://aopkb.org/common/AOP_Handbook.pdf>.
OECD, 2014b. Organisation for economic co-operation and development. Guidance on Grouping of Chemicals: Second Edition. Environment Directorate. (Series on Testing and Assessment No.194). Report No.: ENV/JM/MONO(2014)4, JT03356214. Paris (FR): OECD.
OECD, 2014c. Guidance document for describing non-guideline in vitro test methods. Series on Testing and Assessment No. 211. Available at: <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO%282014%2935&doclanguage=en>.
Perkins, E.J., Antczak, P., Burgoon, L., Falciani, F., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M., Willett, C. In preparation. Quantitative Adverse Outcome Pathways for Regulatory Applications. <https://aopkb.org/saop/workshops/somma.html>.
Petersen, K., Tollefsen, K.E., 2012. Combined effects of oestrogen receptor antagonists on in vitro vitellogenesis. Aquat. Toxicol., 46–53
Rotroff, D.M., Dix, D.J., Houck, K.A., Knudsen, T.B., Martin, M.T., McLaurin, K.W., Reif, D.M., Crofton, K.M., Singh, A.V., Xia, M., Huang, R., Judson, R.S., 2013. Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. Environ. Health Perspect. 121 (1), 7–14.
Rhomberg, L.R., Goodman, J.E., Bailey, L.A., Prueitt, R.L., Beck, N.B., Bevan, C., Honeycutt, M., Kaminski, N.E., Paoli, G., Pottenger, L.H., Scherer, R.W., Wise, K.C., Becker, R.A., 2013. A survey of frameworks for best practices in weight-of-evidence analyses. Crit. Rev. Toxicol. 43, 753–784.
Skolness, S.Y., Blanksma, C.A., Cavallin, J.E., Churchill, J.J., Durhan, E.J., Jensen, K.M., Johnson, R.D., Kahl, M.D., Makynen, E.A., Villeneuve, D.L., Ankley, G.T., 2013. Propiconazole inhibits steroidogenesis and reproduction in the fathead minnow (Pimephales promelas). Toxicol. Sci. 132, 284–297.
Sumpter, J.P., Jobling, S., 1995. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ. Health Perspect. 103 (Suppl 7), 173–178.
Sun, L., Zha, J., Spear, P.A., Wang, Z., 2007. Toxicity of the aromatase inhibitor letrozole to Japanese medaka (Oryzias latipes) eggs, larvae and breeding adults. Comp. Biochem. Physiol. 145C, 533–541.
Sun, L., Wen, L., Shao, Z., Qian, H., Jin, Y., Liu, W., Fu, Z., 2010. Screening of chemicals with anti-estrogenic activity using in vitro and in vivo vitellogenin induction responses in zebrafish (Danio rerio). Chemosphere 78, 793–799.
Tollefsen, K.E., Scholz, S., Cronin, M.T., Edwards, S.W., de Knecht, J., Crofton, K., Garcia-Reyero, N., Hartung, T., Worth, A., Patlewicz, G., 2014. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). Regul. Toxicol. Pharmacol. 70, 629–640.
Villeneuve, D.L., Mueller, N.D., Martinovic, D., Makynen, E.A., Kahl, M.D., Jensen, K.M., Durhan, E.J., Cavallin, J.E., Bencic, D., Ankley, G.T., 2009. Direct effects, compensation, and recovery in female fathead minnows exposed to a model aromatase inhibitor. Environ. Health Perspect. 117, 624–631.
Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M., 2014a. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci. 142 (2), 312–320. http://dx.doi.org/10.1093/toxsci/kfu199.
Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M., 2014b. Adverse outcome pathway development II: best practices. Toxicol Sci. 142 (2), 321–330.
Weed, D.L., 2005. Weight of evidence: a review of concept and methods. Risk Anal. 25, 1545–1557.
Yauk, C.L., Lambert, I.B., Meek, M.E. (Bette), Douglas, G.R., Marchetti, F., Submitted for publication. Development of the Adverse Outcome Pathway 'alkylation of DNA in pre-meiotic male germ cells leading to heritable mutations' using the OECD's Users' Handbook Supplement. Envir. Mol. Mutagen. http://dx.doi.org/10.1002/em.21954.
Yu, L., Liu, C., Chen, Q., Zhou, B., 2014. Endocrine disruption and reproduction impairment in zebrafish after long-term exposure to DE-71. Environ. Toxicol. Chem. 33 (6), 1354–1362.

Annex: Case Examples

List of abbreviations for Annex

TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
AO	adverse outcome
AOPs	adverse outcome pathways
ARNT	AhR nuclear translocator
AhR	arylhydrocarbon receptor
CYP	cytochrome P450
ER	estrogen receptor
ERα	estrogen receptor α
CrVI	hexavalent chromium
JH	juvenile hormone
KE	key event
KERs	key event relationships
Met	methoprene-tolerant
MOA	modes of action
MIE	molecular initiating event
NADPH	nicotinamide adenine dinucleotide phosphate, reduced
NTP	National Toxicology Program
NTE	neuropathy target esterase
OPIDN	OP-induced delayed neuropathy
OECD	Organization for Economic Co-operation and Development
OPs	organophosphates
PHAHs	polyhalogenated aromatic hydrocarbons
PAHs	polyaromatic hydrocarbons
ROS	reaction oxygen species
SDD	sodium dichromate dihydrate
SRC	steroid receptor coactivator
TA	transit amplifying
TOCP	tri-ortho-cresyl phosphate
US EPA	US Environmental Protection Agency
WoE	weight of evidence
XRE	xenobiotic response elements

A.1. Case example: arylhydrocarbon receptor (AhR) activation (MIE) leading to induction of cytochrome P450 CYP) monooxygenases (KE1) and oxidative stress (KE2): evaluation of a subset of KEs and KERs

A number of xenobiotics, including polyaromatic halogenated hydrocarbons (PHAHs) such as 2,3,7,8-tetrachlorodiben

zo-p-dioxin (TCDD), and polyaromatic hydrocarbons (PAHs) such as benzo(y)pyrene, act as agonists of the arylhydrocarbon receptor (AhR). The ligand–receptor complex translocates to the nucleus and dimerizes with the AhR nuclear translocator (ARNT). The AhR/ARNT dimer binds to xenobiotic response elements (XRE) of genes and thereby activates their transcription. The AhR pathway regulates a battery of genes, including several cytochrome P450 monooxygenases, notably those of the CYP1A family (CYP1A1, CYP1A2, CYP1B1). Xenobiotic activation of the AhR pathway can be associated with the production of oxidative stress which is considered to play a major role in the toxicity of AhR-binding chemicals, in particular in causing mutagenicity and tumour formation. The term oxidative stress refers to any shift of the cellular redox homeostasis towards the increased production of reaction oxygen species (ROS) relative to the cellular antioxidant defence.

Here, we discuss an AOP linking AhR activation to tumor formation. For illustrative purposes (Fig. A1), we focus only on the following steps: the MIE which is xenobiotic binding to and activation of the AhR, the subsequent induction of CYP enzymes (KE1), and the induction of an oxidative stress response (KE2).

A.1.1. Biological plausibility (AhR induction of cytochrome CYP monooxygenases leading to oxidative stress)

There exists a well-confirmed mechanistic relationship between the MIE (AhR activation) and KE1 (CYP monooxygenase induction) (Beischlag et al., 2008). A plausible mechanism exists also for the induction of oxidative stress by CYP monooxygenase activity. In the course of the reaction catalyzed by CYP monooxygenases, two electrons are sequentially transferred from NADPH-dependent cytochrome P450 oxidoreductase to each atom of bound oxygen, resulting in the production of oxygenated substrate and water. Although tight coupling normally exists between oxygen reduction and monooxygenation, some reactive oxygen may be released as either superoxide or H_2O_2 in the course of the electron transfer. In this way, CYP monooxygenases can produce ROS (Dalton et al., 2002; Reichard et al., 2006), with the rate of CYP-dependent ROS generation depending on the efficiency – or inefficiency – of the coupling of NADPH consumption to substrate oxidation (Schlezinger et al., 1999; Zangar et al., 2004). In fact, the ability of CYP-containing microsomes for NADPH-dependent generation of ROS has been known for over 50 years (Gillette et al., 1957).

Importantly, however, mechanisms other than monooxygenase-dependent ROS production can lead from AhR activation to oxidative stress. One possible mechanism is that CYP1A-catalyzed metabolism produces oxidative stress not by direct ROS generation, but indirectly via generation of reactive metabolites. For instance, it is known that female rats are more susceptible to TCDD-induced oxidative stress than males (Dalton et al., 2002). The causative mechanism involves, at least in part, monooxygenase-catalyzed metabolism of estrogen. Since TCDD affects the activity of CYP monooxygenases, it can thereby alter the production and the ratios of electrophilic metabolites of estradiol such as the hydroxy-E2 catechols, which in turn alter the cellular oxidation status (Dalton et al., 2002). Another possible mechanism is the AhR-regulated reduction of the cellular antioxidant capacity, for instance, by induction of cytoprotective genes such as NAD(P)H:quinone oxidoreductase 1, glutathione-S-transferases and UDP glucuronosyltransferases, which combat oxidative stress (Latchoumycandane et al., 2002; Kalthoff et al., 2010). This latter mechanism involves a crosstalk between the AhR and another transcription factor, the nuclear factor erythroid 2-related factor (nrf2) (Lu et al., 2011). Another CYP-independent mechanism of oxidative stress production by AhR ligands seems to be the induction of cytokines (Nebert et al., 2000; Tsuji et al., 2011). Finally, there exists evidence that dioxins can induce mitochondrial ROS production by a mechanism independent of CYP monooxygenase activity (Senft et al., 2002). Thus, in all likelihood, there are several AhR-dependent pathways that lead to oxidative stress, involving both an increase of monooxygenase-dependent ROS production as well as other pathways (Dalton et al., 2002).

A.1.2. Essentiality (AhR induction of CYP monooxygenases leading to oxidative stress)

Evidence shows that KE1 can be prevented by blockage of AhR binding as well as in AhR knockout animals; thus, AhR activation is essential for induction of monooxygenases such as CYP1A (Beischlag et al., 2008). The situation is less clear for the relation between CYP monooxygenase activity and ROS production. Kopf et al. (2010) showed that CYP1A knockout mice accumulated TCDD at similar levels as wild type mice, but in contrast to wild type mice, did not show ROS production. Related to this is the observation that toxicological sensitivity to TCDD differs between $ahr^{-/-}$ and $cyp1a^{-/-}$ mice (Reichard et al., 2006); knockout of CYP1A protected the mice against TCDD doses up to 200 μg/kg, whereas AhR knockout protected against TCDD doses up to 2 mg/kg. Another interesting observation concerning the relationship between AhR, CYP monooxygenases and ROS is that TCDD-exposed $cyp1a^{-/-}$ could produce mitochondrial ROS despite the absence of CYP1A. This finding supports the hypothesis that CYP1A induction is not essential for AhR-dependent ROS production in the mitochondria. Furthermore, the finding that $ahr^{-/-}$ mice were not capable of mitochondrial ROS production after TCDD exposure (Senft et al., 2002) points to the essentiality of AhR activation. Overall, the available data, although fragmented, point to an essential role of AhR in ROS generation and oxidative stress, while the essentiality of CYP1A appears to be moderate at best.

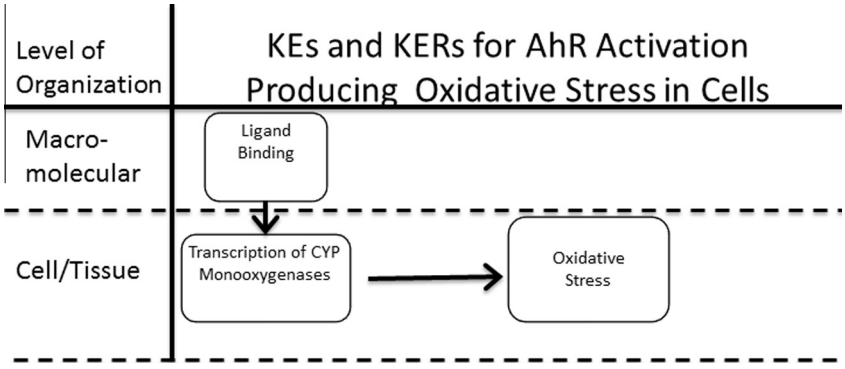

Fig. A1. Depiction of a subset of KEs and KERs relating arylhydrocarbon receptor (AhR) activation to the induction of cytochrome P450 (CYP) monooxygenases and oxidative stress.

A.1.3. Empirical support (AhR induction of CYP monooxygenases leading to oxidative stress)

Numerous studies report that activation of the AhR pathway is associated with oxidative stress. As shown by Schlezinger and Stegeman (2001) for the teleost species, *Stenotomus chrysops*, there exists a significant correlation between the level of microsomal CYP1A activity and the microsomal rate of ROS production. Pronounced and prolonged induction of oxidative stress has been shown to occur in humans and rodents after exposure to AhR-binding PHAHs and PAHs (Shertzer et al., 1998; Slezak et al., 2000; Dalton et al., 2002; Reichard et al., 2006; Costa et al., 2010; Tsuji et al., 2011). ROS production after TCDD exposure is significantly higher in C57BL/6 mice, which carry the high-affinity ahr^{b1} allele, compared to the low-affinity DBA/2 mice (Alsharif et al., 1994). Induction of oxidative stress by AhR-binding xenobiotics has also been shown in fish (Palace et al., 1996; Schlezinger and Stegeman, 2001). Dose- and time-dependent relationships between AhR ligand exposure and ROS production can be complex: for instance, in mice, higher tissue concentrations of TCDD were required to elicit oxidative stress responses following acute exposure than with subchonic exposure (Reichard et al., 2006). This may relate to adaptive/compensatory processes which influence dose– as well as time–response relationships so that in many cases $KE_{upstream}$ does not linearly translate into $KE_{downstream}$.

The pathway from the MIE to oxidative stress involves two KERs. For KER1, extensive empirical evidence supports a direct, both dose- and time-dependent linkage between AhR activation and CYP monooxygenase gene expression (Beischlag et al., 2008). Also for KER2, good empirical evidence exists for different animal taxa that elevated CYP monooxygenase activity associates with oxidative stress (Dalton et al., 2002; Curtis et al., 2011). A well-studied example that AhR-mediated induction of CYP monooxygenases results in the production of oxidative stress comes from studies on the effect of dioxins on vascular endothelia (Kopf and Walker, 2010).

A.1.4. WoE conclusions (AhR induction of cytochrome P450 enzymes leading to oxidative stress)

Based on considerations described above, the WoE for KER1 is strong. The relation between KE1 and KE2, however, is more equivocal. There is clear evidence that in many cases AhR-mediated increase of CYP monooxygenase activity (in particular of CYP1A) results via increased ROS production in the production of oxidative stress; however, there also exist AhR-regulated but CYP-independent pathways of oxidative stress production. The situation becomes even more complicated if we do not consider the relation between CYP and ROS, but between CYP and oxidative stress. AhR-induced oxidative stress does not necessarily occur via CYP monooxygenase activity and CYP-dependent ROS production, but there exist alternative mechanisms linking AhR activation to oxidative stress. Thus, based on considerations of essentiality and empirical evidence, WoE for KER2 is only moderate.

A.2. Case example: juvenile hormone agonist-induction leading to increase in male offspring in the arthropod cladocera: evaluation of a subset of KEs and KERs

In the cladocera, commonly called water flea including the *Daphnia* species, juvenile hormone (JH) regulates important physiological and developmental processes, such as molting, growth, reproduction, and sex determination. Insects and other crustaceans also use JHs with similar structure and various JH analogs (agonists) have been developed for the control of insect growth. JH analogs, such as pyriproxyfen, fenoxycarb and diofenolan, and also known as insect growth regulators, have been reported to induce male offspring in the cladocera. The cladocera generally reproduce female offspring by parthenogenesis; however, when the environmental conditions worsen (short day length, food shortage, an increase in population density), they produce male offspring and sexual reproduction occurs. Male offspring production by topical application of crustacean JH (methyl farnesoate), insect JH (JHIII) and JH analogs occurs independent of environmental conditions, suggesting that JH is a key endocrine factor for sex differentiation working downstream of environmental stimuli in the cladocera (Olmstead and LeBlanc, 2003; Tatarazako et al., 2003; Oda et al., 2005; Abe et al., 2015). Therefore, identification sex of offspring by the length of first antenna was added as the optional endpoint in *D. magna* reproduction test (OECD TG 211, annex 7) to evaluate JH action.

A pathway of sex determination/differentiation regulated by JH has been gradually uncovered mainly in *Daphnia* species. As a MIE, JH and its analogs interact with the JH receptor (a heterodimer of Methoprene-tolerant (Met) and steroid receptor coactivator (SRC) proteins (Miyakawa et al., 2013; LeBlanc et al., 2013). Activation of the JH receptor with JH or JH agonists most likely induces *Doublesex1* gene expression (KE1) later on during ovarian development, resulting in the male production (KE2) (Kato et al., 2011). Increase of male population may lead to reduction of reproductive rate in the population (AO) because sexual reproduction produces only two female offspring from dominant eggs per brood (Fig. A2).

Currently, insufficient direct evidence of the KER between JH receptor binding by JH and its analogs (MIE) and activation of *Doublesex1* gene expression (KE1) is available since the *Doublesex1* gene has only been recently identified (Kato et al., 2011). Therefore, we evaluate the indirect relationship between JH receptor binding (MIE) and the increase in production of male offspring (KE2) for purposes of the WoE analysis.

A.2.1. Biological plausibility (JH receptor binding by an agonist (MIE) and increase in male offspring (KE2))

While several JH agonists (e.g. fenoxycarb, pyriproxyfen, methoprene, epofenonane and diofenolan), which induce male offspring in several species of the cladocera in *in vivo* tests were reported to bind to the JH receptor (Met and SRC Heterodimer), the understanding of the downstream KEs after key gene expression (KE1) that result in the production of male offspring is not complete, in particular KEs at the cellular and organ level. Thus, the biological plausibility between JH receptor binding by agonists and an increase in male offspring production is considered moderate.

A.2.2. Essentiality (JH receptor binding by an agonist (MIE) and increase in male offspring (KE2))

Essentiality of JH receptor binding is considered strong because there is direct experimental evidence showing that transcriptional knockdown of the JH receptor gene in *Daphnia magna* embryos resulted in embryonic death (Miyakawa et al., 2013). In addition, transcriptional knockdown of *Doublesex1* gene in the male embryo which was induced by exposure to fenoxycarb did not develop the male phenotype, suggesting that *Doublesex1* gene expression (KE1) is essential for male offspring development (Kato et al., 2011).

A.2.3. Empirical support (JH receptor binding by an agonist (MIE) and increase in male offspring (KE2))

JH receptor binding studies in *D. magna* and *D. pulex* which use a two-hybrid assay (Miyakawa et al., 2013) and luciferase reporter gene assays (LeBlanc et al., 2013) showed that transcriptional activation of JH receptor by JH agonists *in vitro* occurs at lower concentrations than those inducing male offspring *in vivo* (Tatarazako et al., 2003). With respect to temporal concordance, production of male offspring is

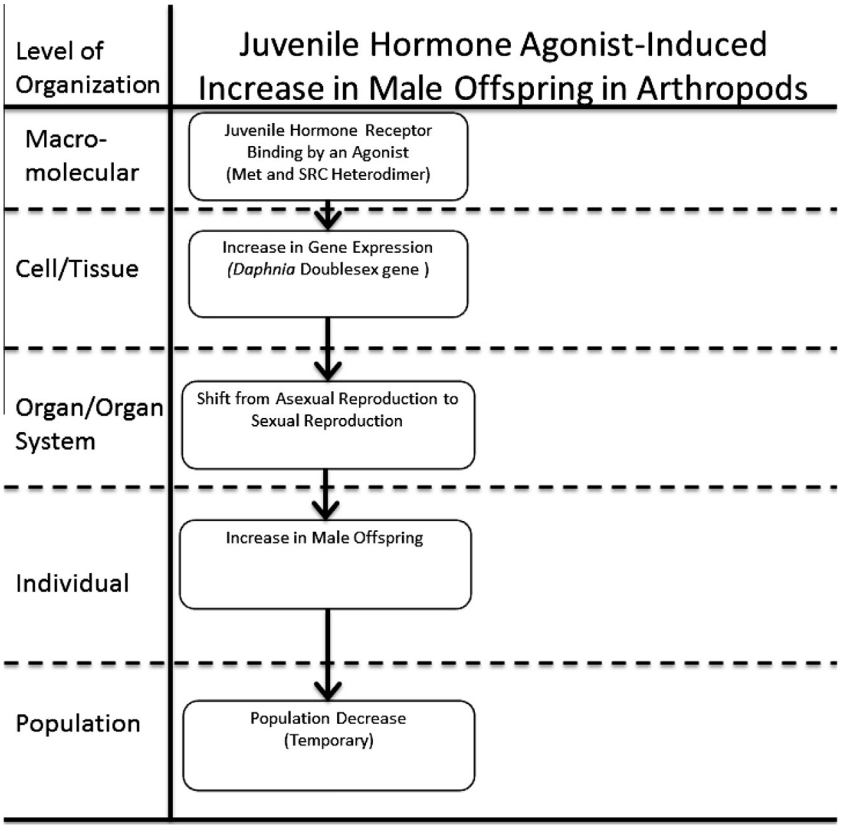

Fig. A2. Depiction of the subset of KEs and KERs in the juvenile hormone agonist-induced increase in male offspring in the arthropod cladocera AOP.

induced by JH and JH analogs in the critical period (i.e. 7–8 h before ovulation) (Kato et al., 2010), and expression of *Doublesex1* gene is up-regulated 18 h after ovulation in JH-exposed offspring (Kato et al., 2011). This indicates that JH receptor binding (MIE), which is probably followed by *Doublesex1* gene expression (KE1), occurs in the first half period of embryo development in the brood chamber after exposure of JH and JH analogs. Regarding the evaluation of consistency, male offspring production by JH agonists was observed in six cladoceran species (Olmstead and LeBlanc, 2003; Toyota et al., 2013). Accordingly, consistency across species is considered high (strong). However, Miyakawa et al. (2013), reported a juvenoid (e.g. epofenonane) induced male offspring without interaction of Met and SRC heterodimer JH receptor. This suggests that there may be interactions with other receptor(s) or other steps in the pathway (e.g. stimulation of JH secretion) that are not yet well understood. As the focus of this evaluation was on the liner pathway between JH receptor binding and male offspring production, this result is not considered empirical evidence for inconsistencies of the KER. Therefore, empirical support for the KER between MIE and KE2 is considered strong.

A.2.4. WoE conclusions (JH receptor binding by an agonist (MIE) and increase in male offspring (KE2))

Based on moderate biological plausibility and strong essentiality and empirical support, overall WoE for the indirect KER between JH receptor binding by an agonist and increase in male offspring is considered strong.

A.3. Case example: binding of certain organophosphates to neuropathy target esterase (NTE) leading to delayed neuropathy: evaluation of a subset of KEs and KERs

A well-known chemically induced neuropathy is a central/peripheral sensory-motor distal axonopathy that appears up to several weeks after exposure to certain organophosphates (OPs), named OP-induced delayed neuropathy (OPIDN) (Abou Donia and Lapadula, 1990; Weiner and Jortner, 1999). It is clear that not all OPs can cause OPIDN. Selected OPs used in aviation fluids or as oil additives such as tri-ortho-cresyl phosphate (TOCP) and insecticides like chlorpyrifos, dichlorvos, isofenphos, methamidophos, mipafox and trichlorfon have been shown to induce OPIDN in humans and animal models (Weiner and Jortner, 1999; Lotti and Moretto, 2005). These OPIDN inducers covalently bind to the active site of neuropathy target esterase (NTE) (Johnson, 1990), causing not only inhibition but also "aging" of the enzyme, a process that is characterized by loss of an R-group from the phosphoryl moiety that leads to the formation of a negatively charged phosphoryl group, which is covalently bonded to the active site serine of the esterase (Richardson et al., 2013). The majority of experimental data derives from adult hens (18 weeks old), the animal model of choice for the identification of OPIDN inducers (Doherty, 2006). On the other hand, adult mice seem to be resistant to similar insults (Veronesi et al., 1991) and present axonal degeneration and paralysis only after long exposure up to 9 months to these compounds (Lapadula et al., 1985).

This AOP has as a MIE the binding of OPIDN inducers to NTE, which is followed by the cellular KEs: NTE inhibition and "aging", disruption of Ca^{2+} homeostasis, mitochondrial dysfunction and disruption of neuronal cytoskeleton. These cellular effects result in peripheral sensory and motor distal axonopathy that is manifested as peripheral neuropathy (Fig. A3).

Here, for illustrative purposes, we explore two different types of pairs of KEs from this AOP in an attempt to apply the evolved WoE considerations based on Table 1. More specifically, the relationship between inhibition and "aging" of NTE and increase of intracellular Ca^{2+} (KER1), as well as increase of intracellular Ca^{2+} and mitochondrial dysfunction (KER2) are discussed.

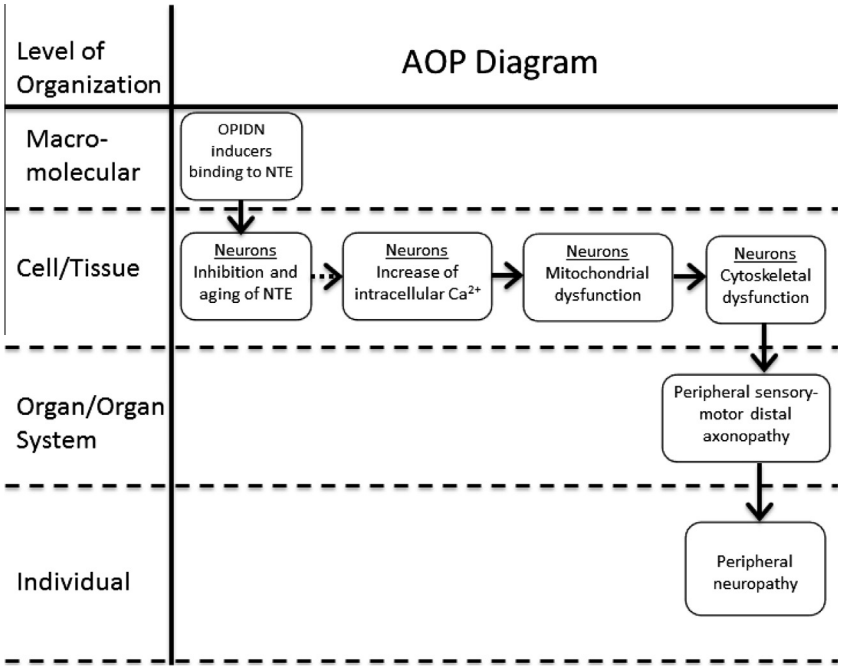

Fig. A3. Depiction of the AOP for binding of certain organophosphates to neuropathy target esterase (NTE) leading to development of delayed peripheral neuropathy.

A.3.1. KER1 (the relationship between inhibition and "aging" of NTE and increase of intracellular Ca^{2+})

A.3.1.1. Biological plausibility of KER1

The biological plausibility of the relationship between inhibition and "aging" of NTE and increase of intracellular Ca^{2+} (KER1) is low (weak). While there is considerable understanding of the molecular interactions and pathways involved in OP delayed neurotoxicity, the exact mechanism that would explain how inhibition and "aging" of NTE indirectly leads to disruption of Ca^{2+} homeostasis is not known. Several *in vitro* and *ex vivo* studies using brain tissue from treated hens support that OPIDN inducers can "age" NTE, whereas other OPs are capable of binding to the active site of NTE and inhibiting its activity without causing "aging", and consequently, without producing OPIDN (reviewed in Hargreaves, 2012). However, the events that occur between NTE "aging" and the manifestation of clinical symptoms that characterize OPIDN are not fully understood. It has been suggested that NTE "aging" is followed by an increase in intracellular Ca^{2+} that affects neuronal cytoskeleton leading to axonal degeneration (reviewed in Emerick et al., 2012).

A.3.1.2. Essentiality of KER1

With respect to essentiality, only indirect evidence exists to support this KER. This indirect evidence shows that when Ca^{2+} channel blockers are administered to animals prior to OP exposure there is a reduction of OPIDN symptoms and most importantly, the same blockers given after certain OP exposures significantly protect and alleviate the clinical symptoms (Emerick et al., 2012). In an experimental study, co-exposure of rats to a Ca^{2+} channel blocker and dichlorvos prevented neurobehavioral alterations induced by dichlorvos but no prevention of NTE inhibition has been found (Choudhary and Gill, 2001). Although the essentiality of this KE (increase of intracellular Ca^{2+}) is strong for all the downstream KEs and AO, the same is not valid for the relationship between NTE aging and the increased intracellular Ca^{2+} levels (Bal-Price et al., 2015).

A.3.1.3. Empirical support of KER1

Experiments have yet to be conducted to directly address dose–response relationships between these two KEs. Therefore, only indirect evidence can be used to evaluate WoE for this KER. The extensive set of studies reviewed in Emerick et al. (2012) to assess the reversibility and protection by Ca^{2+} channel blockers at the organ and organism levels used several OPIDN inducers at concentrations that are known to cause NTE inhibition and "aging". With respect to temporal concordance, the NTE inhibition appears in a much earlier time point than the increase in intracellular Ca^{2+} levels. For example, the administration of dichlorvos to rats showed brain NTE inhibition up to 65.2% compared to controls after 24 h and recovery of the enzyme up to 84% before the manifestation of clinical signs (day 21), whereas the maximum increase (2.74-fold) in free ionic Ca^{2+} in the cytosol was observed on day 15 and kept high with slight drop until day 21 (Choudhary and Gill, 2001). In terms of consistency across species, NTE is highly conserved across species including mammals, insects, nematodes, and yeast (Moser et al., 2000). The same is observed for the Ca^{2+} homeostatic system that is known to be highly conserved throughout evolution (Case et al., 2007). Hens have been mainly used in studies of OPIDN, however, rats have also been demonstrated to be a valid model showing susceptibility and NTE inhibition but no appearance of locomotor ataxia that is evident in hens and humans after exposure to OPIDN inducers. It is suggested that these differences in locomotor ataxia may be due to differences in toxicokinetics and more rapid recovery of NTE in the rat brain compared to other species. Regarding the evaluation of consistency based on empirical data, there is high degree of uncertainty of the relationship between inhibition and "aging" of NTE and disruption of Ca^{2+} homeostasis, which derives mainly from the lack of knowledge on the mechanism/s involved. Consequently, through an overall WoE evaluation, the linkage between inhibition and "aging" of NTE and disruption of Ca^{2+} homeostasis (KER1) is considered weak.

A.3.2. KER2 (increase of intracellular Ca^{2+} and mitochondrial dysfunction)

A.3.2.1. Biological plausibility

There is substantial mechanistic understanding (Giorgi et al., 2008) about the interrelationship between disruption of Ca^{2+} homeostasis and mitochondrial dysfunction; therefore biological plausibility is considered to be strong. This interdependence of the two KEs is highly important for cells with elevated energy demands such as neuronal cells. Although there are several cellular organelles that regulate Ca^{2+} homeostasis, mitochondria are of special importance compared to the others as they can initially contribute to the uptake of excessive Ca^{2+} and later to the release of this ion under certain conditions (Giorgi et al., 2008).

A.3.2.2. Essentiality

The essentiality of the disruption of Ca^{2+} homeostasis for this AOP is rated strong as it has been already discussed above. In contrast, no studies directly assessing the essentiality of mitochondrial dysfunction in relation to the AO are available, whereas there is limited experimental evidence supporting the essentiality of this KE in relation to downstream cellular KE cytoskeletal dysfunction (reviewed in Hargreaves, 2012), suggesting that this KE can be rated as moderate.

A.3.2.3. Empirical support

Regarding temporality, the understanding of the relationship between disruption of Ca^{2+} homeostasis and mitochondrial dysfunction (KER2) is considered to be moderate, because just as collapse of Ca^{2+} homeostasis can lead to mitochondrial dysfunction, mitochondrial dysfunction can also contribute to increased intracellular Ca^{2+}. Morphological changes in mitochondria have been shown to appear early (day 5) following exposure of hens to TOCP and moreover, these alterations further developed in a time-dependent manner leading to neuronal loss (Mou et al., 2006) however empirical data is lacking – no measurement of Ca^{2+} has been performed. Thus, direct empirical support is lacking at this time.

In conclusion, although the biological plausibility of the linkage between Ca^{2+} homeostasis and mitochondrial dysfunction for this AOP is strong, in the absence of empirical data and due to the lack of established temporality between the two KEs, the overall WoE for KER2 is considered to be moderate.

A.3.3. WoE conclusions

The WoE of KER1 (the linkage between inhibition and "aging" of NTE and disruption of Ca^{2+} homeostasis) is considered weak. The WoE for KER2 (increase of intracellular Ca^{2+} leading to mitochondrial dysfunction) is considered to be moderate.

A.4. Case example: agonist binding to estrogen receptor α leading to an increased risk of endometrial cancer: evaluation of a subset of KEs and KERs

Estrogens are important regulators of development and functioning of the reproductive system in female, but also male higher vertebrates. They also influence many physiological processes not strictly related to reproduction, including cardiovascular health, bone physiology, cognition, and behavior (reviewed by Deroo and Korach, 2006; Heldring et al., 2007). Aberrant exposure to estrogens can lead to diseases in target organs. For instance use of estrogens, without addition of synthetic progesterone compounds, can lead to an increased risk of endometrial cancer, on the order of 2–12-fold; use for 5–10 years appears to lead to the greatest risk (http://www.drugs.com/pro/estradiol.html#ID_690ebdc9-1c4f-843a-e6a8-1e21f7aa7da9). The pathways of estrogen action in stimulating endometrial growth have been well characterized, with an early key event being transcription of target genes (see for example, Gielen, 2005; citebib26; Heldring et al., 2007). Such sustained exposure to estrogens, in the absence of the addition of progesterone agents, can lead to endometrial hyperplasia, which may be a precursor to endometrial cancer. Using an animal model, Moggs et al., 2004 reports estrogen effects on the endometrium, specifically noting that it "...begins with the induction of genes involved in transcriptional regulation and signal transduction and is followed, sequentially, by the regulation of genes involved in protein biosynthesis, cell proliferation, and epithelial cell differentiation. Furthermore, we have identified genes with common molecular functions that may drive fluid uptake, coordinated cell division, and remodeling of luminal epithelial cells."

For illustrative purposes, we focus on one KER in the AOP; the KER between estrogen receptor α (ERα) binding by an agonist (MIE) and transcriptional activation (KE1) (Fig. A4). A wide number of transcripts could potentially be evaluated, as indicated in Moggs et al., 2004. In practice, transcriptional activation of ERα can be accurately assessed via reporter genes containing estrogen responsive elements as the sole responsive elements (Sonneveld et al., 2006). Test guidelines have been formulated that can be updated regularly, describing detailed procedures to carry out such reporter gene assays (OECD, 2009, 2012; EDSP, 2009). Other possibilities to measure transcriptional activation through estrogen receptor α include measurement of ligand-induced expression of endogenous target genes that are expressed in target cells. This can have practical advantages showing responses in intact organisms allowing linkage of estrogen receptor activation to adversity, but also has drawbacks with respect to specificity and interpretation of the results. The main problem with respect to specificity is the possibility of regulation by other transcription factors of these target genes, of which some factors may be influenced by the chemical which is under investigation as the potential ERα activating compound. In addition, because of overlapping activities it can be difficult to distinguish between ERα and ERβ activating compounds (Kuiper et al., 1998). Therefore, it may be preferable to couple testing using less selective tests with methods that use a more selective test (i.e. a selective, validated reporter gene assay), or a specificity control using a receptor antagonist.

A.4.1. Biological plausibility (KER between estrogen receptor α binding by an agonist (MIE) and transcriptional activation (KE1))

The main endogenous physiological estrogen is 17β-estradiol, while strong synthetic ligands such as diethylstilbestrol have been designed for pharmaceutical applications. In addition, many ligands are known from a variety of sources, often much weaker than aforementioned ones, including phytoestrogens and chemicals with unintended weak estrogenic action that have been linked to endocrine disruption (McLachlan, 2001). Some ligands can act as antagonists, diminishing or even blocking the action of agonistic ligands (Deroo and Korach, 2006; Heldring et al., 2007). Most of the effects of estrogens are mediated through the estrogen receptors (ERs), ERα (NR3A1) and ERβ (NR3A2), conserved proteins evolving from a common vertebrate ancestor (Thornton, 2001). These receptors belong to the nuclear hormone receptor family and have distinct domains that have different functions, such as ligand binding as well as those mediating DNA binding and transcriptional activation upon ligand binding. Modulation of transcriptional activity can occur through cross-talk with other signaling pathways through phosphorylation, or protein–protein interactions. In addition, some estrogen effects are mediated through non-genomic actions as a result of signaling by estrogen

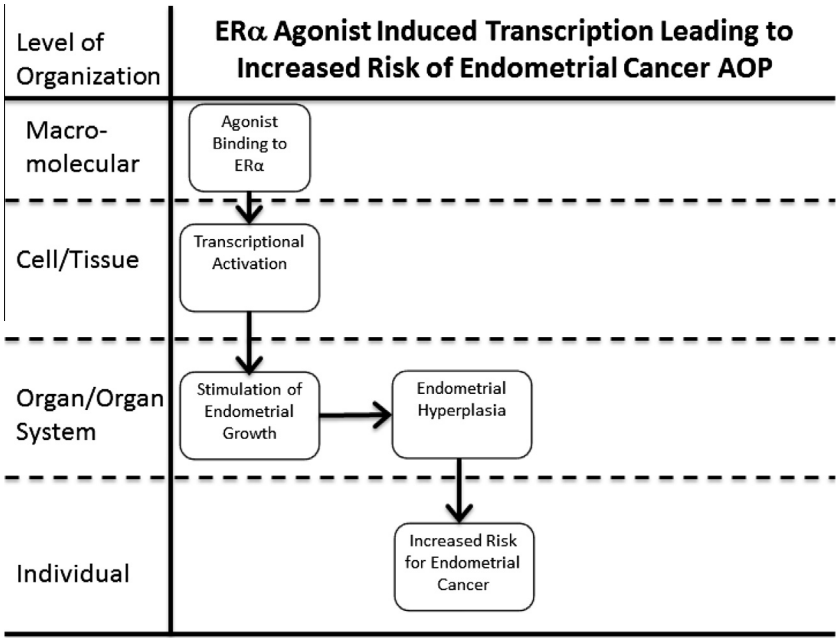

Fig. A4. Depiction of the AOP of agonist binding to estrogen receptor α leading to an increased risk of endometrial cancer.

receptors bound through adapter proteins to membrane receptors (Deroo and Korach, 2006; Heldring et al., 2007). There are few uncertainties and there is little conflicting evidence with respect to the relationship of agonist binding to ERα and transcriptional activation. The KER relationship is highly conserved in mammals and higher vertebrates, and there are many well documented examples. The estrogen receptor is expressed in all vertebrates and even classes of invertebrates. Evidence suggests, however, that some invertebrate and low vertebrate ER do not bind 17β-estradiol, showing that care should be taken in extrapolating this KER when considering effects in lower species (Paris et al., 2008).

The biological plausibility of the KER between ERα binding (MIE) and subsequent transcriptional activation (KE1) by agonistic ligands is considered high (strong). An extensive body of scientific evidence has accumulated documenting the pathway linking ERα agonist binding and induction of transcription. In brief, there is a direct link between binding of receptor agonists and transcriptional activation. The transcriptional activation of target genes is a result of binding of the receptor to specific binding sites (estrogen responsive elements) in promoter regions of these genes, and transcription occurs through a series of steps that include co-regulator recruitment and through this increased transcription of target genes by the basic transcriptional machinery. These steps have been analyzed in detail and described extensively in the literature. Ligand-independent methods of activation have also been described, that involve phosphorylation of the receptor via pathways activated by polypeptide growth factors. However, these seem not to be involved in the majority of physiological functions and do not play a significant role in the KER interaction when assessing the effect of ligands under well-defined experimental conditions, with constant levels of (serum-derived) polypeptide growth factors.

The above strong linkage of transcriptional activation upon ligand binding, however, does not apply to antagonists. These are a subset of receptor binding compounds, that do not (in the case of full antagonists), or only partially (in the case of partial antagonists), activate the receptor since the conformational change upon receptor binding is such that no or different transcriptional co-regulators are recruited. Essentiality of the relationship can be further distorted if there is different metabolic capacity in the test systems assessing receptor binding and transcriptional activation. This can lead to metabolites with either stronger receptor binding and transcriptional activation (e.g. in the case of methoxychlor) or weaker activities (e.g. in the case of bisphenol A). In the case of equal or absence of metabolic capacity in test systems this consideration does not apply. Again, this shows the importance of using well-defined experimental systems and conditions in assessing ERα activation.

A.4.2. Essentiality (KER between estrogen receptor α binding by an agonist (MIE) and transcriptional activation (KE1))

The support for essentiality of this KER is considered high (strong). Essentiality is clearly demonstrated in studies that use antagonists (e.g. Tremblay et al., 1998). In addition, estrogen effects are clearly linked to expression of estrogen receptors and mutations in the receptor that prevent ligand binding also prevent transcriptional activation. Furthermore, ER knock-out animals show strongly impaired reproductive functions (Deroo and Korach, 2006; Heldring et al., 2007)

A.4.3. Empirical support (KER between estrogen receptor α binding by an agonist (MIE) and transcriptional activation (KE1))

The empirical support of KERs is considered high (strong). There is extensive detailed understanding of the KER, including the sequence and the manner in which ligands bind to and alter the conformation of ERαs, the binding of the ligand–receptor complex to DNA, recruitment of transcriptional co-regulators and subsequent transcriptional activation. Extensive dose–response relationships with multiple chemicals are available showing concordance between receptor binding and transcriptional activation (Sonneveld et al., 2006). Transcriptional activation can be measured with a reporter gene assay. In these assays ERα activation is measured with a reporter gene consisting of the estrogen responsive elements coupled to a gene coding for a protein which can be easily and specifically measured (e.g. luciferase or green fluorescent protein) (Legler et al., 1999).

A.4.4. WoE conclusions

The WoE of KER1 (agonist binding to ERα leading transcription) is high (strong).

A.5. A chemical specific case example: induction of cytotoxicity and regenerative hyperplasia by oral hexavalent chromium (Cr(VI)) leading to duodenal tumors in mice

In this AOP case example, we focus on the tumor promotional mode of action produced by high oral doses of chromium VI (CrVI) in mice. As an AOP case example applicable to a single chemical or a very limited number of chemicals (OECD AOP Project Proposal Guidance (OECD, 2009)), the WoE evaluation is largely limited to the evidence developed for the chemical for which the AOP has been constructed. As discussed in Section 4.6, the role of chemical specific AOPs, as knowledge on this MOA increases, the breadth of application of a chemical-specific case example could expand to a broader chemical domain.

In the intestinal tract as well as other tissues, the multistep process of carcinogenesis is characterized by the canonical phases of initiation, promotion, and progression. The initiation phase is generally considered to result from mutations that may occur spontaneously or through exposure to chemicals. The tumor promotion phase typically requires a sustained cellular growth/cell proliferation stimulus as well as inhibition of apoptosis of initiated cells. These initiated cells thus acquire a selective advantage and undergo clonal expansion to form small foci of altered cells and, eventually, tumors. These foci of altered/initiated cells have been extensively characterized in colonic crypts in rodents and humans where they are clearly visible during histopathological examination (Cheng and Lai, 2003; Takahashi and Wakabayashi, 2004).

Duodenal tumors in mice are believed to originate from mutations in the stem cells that reside at the base of the intestinal crypts. A small number of stem cells located near the base of the crypts of Lieberkühn give rise to proliferating progenitor or transit amplifying (TA) cells (Chia and Kuo, 2010; Shaker and Rubin, 2010). The TA cells occupy the length of the crypts and give rise to eight different types of intestinal cells, including the absorptive enterocytes. Complete turnover of the epithelium occurs every 3–5 days in an orderly fashion along the crypt–villus axis as enterocytes move upward to the villi. TA cells differentiate to become and replace older enterocytes that are sloughed off into the intestinal lumen.

Exposure to certain non-genotoxic agents at sufficiently high doses can cause cytotoxicity to differentiated intestinal villous cells. If the exposure is episodic, a burst of proliferation in the crypts will occur to re-populate the cells lost due to cytotoxicity. However, if the exposure is sustained and produces prolonged cytotoxicity, a correspondingly sustained cell proliferation is induced. In such a situation, such high rates of proliferation create a situation in which insufficient time exists to repair any DNA damage occurring spontaneously or from genotoxic agents. Thus, miss-repair of DNA that leads to mutations also becomes more likely in this sustained proliferative environment.

Both inhibition of apoptosis and the proliferative environment serve to promote the growth of spontaneously initiated cells. Intestinal initiation/promotion models indicate the MOA of induction of adenocarcinomas involves cytotoxicity and regenerative cell proliferation, a common mechanism in the MOA of many tumors (Cohen and Arnold, 2011; Pitot et al., 2000). In the duodenum of the mouse, tumorigenesis can occur "by a non-genotoxic MOA involving cytotoxicity and regenerative cell hyperplasia" and such a MOA has been characterized as exhibiting "a clear dose threshold" (USEPA, 2004). Studies with captan have shown that if the cytotoxic doses are discontinued, the tumor promotion responses are reversible, leading EPA to accept that "there is a strong causal association (dose–response, temporality) indicating that tumor formation is secondary to cytotoxicity and hyperplasia and that the latter is a KE in the sequential cascade of events leading to cancer" (USEPA, 2004). The fungicide folpet, similar in structure to captan, also appears to act via a similar mechanism (Cohen et al., 2010; Gordon et al., 2012)

Duodenal tumors have more recently been observed in mice following oral exposures to CrVI (National Toxicology, 2008). Subsequently, an extensive research program was undertaken to characterize the MOA and dose-dependent sequence of events leading to intestinal tumor formation in mice. Based on these studies the hypothesized sequence of events leading to the AO of duodenal tumors in mice from oral exposures to CrVI in drinking water is depicted as an AOP (Fig. A5).

A.5.1. Biological plausibility (non-genotoxic induction of cytotoxicity and regenerative hyperplasia by a threshold mechanism promotes duodenal tumors in mice)

The WoE for biological plausibility is high (strong). There is textbook level understanding that turnover of normal villi occurs by migration of epithelial cells from the base of the intestinal crypts, and that stem cells within the crypts serve as the source of these new epithelial cells (Johnson, 2013). Similarly, hyperplasia in intestinal crypts is a well-documented response to villous cytotoxicity. Although the theoretical mechanism by which villous cytotoxicity leads to crypt hyperplasia is well established (see Thompson et al., 2013), direct evidence of a resulting proliferative molecular signal, such as a change in cytokine concentration in enterocytes is lacking. Prolonged cytotoxicity associated with regenerative hyperplasia is a well-documented MOA for tumor promotion in a number of tissues for a large range of chemicals in numerous animal models (Boobis et al., 2009; Cohen and Arnold, 2011). This tumor promotion MOA (prolonged exposure at sufficiently high doses leading to sustained tissue injury) has been described in toxicology textbooks since the mid-1980s (e.g. Casarett and Doull's Toxicology, third ed., 1986).

A.5.2. Essentiality

Chronic cytotoxicity in enterocytes in small intestine villi in mice is the MIE. Oral doses that cause such chronic cytotoxicity (e.g. CrVI doses that significantly exceed the gastric reducing capacity for an extensive period of time) cause crypt hyperplasia and eventually tumorigenesis in the mouse intestine. At doses that do not exceed the reducing capacity, cytotoxicity is not observed, and normal cellular homeostasis appears to be maintained (Thompson et al., 2011, 2012, 2013).

Thus, evaluation of essentiality is focused on the KER of cytotoxicity leading to hyperplasia/proliferation. Destruction of villous enterocytes (caused by doses that exceed cellular mechanisms which maintain homeostasis) causes the production of a proliferative signal to the crypt stem cells and resulting hyperplasia in the crypts. Very recent evidence indicates that following exposure of mice to 180 mg/L CrVI in drinking water, chromium occurs close to background levels within cells in the crypt whereas the chromium levels in villous cells were reported to be more than thirty fold higher (Thompson et al., 2015). This finding supports the idea that crypt hyperplasia occurs in response to villous cytotoxicity.

The evidence for essentially is moderate. In the mouse, hyperplasia does not occur at doses below those that cause villous enterocyte cytotoxicity. Evidence of essentiality is supported by the NTP studies in rats; specifically, villous blunting indicative of cytotoxicity did not occur in rats treated with oral doses of CrVI) that

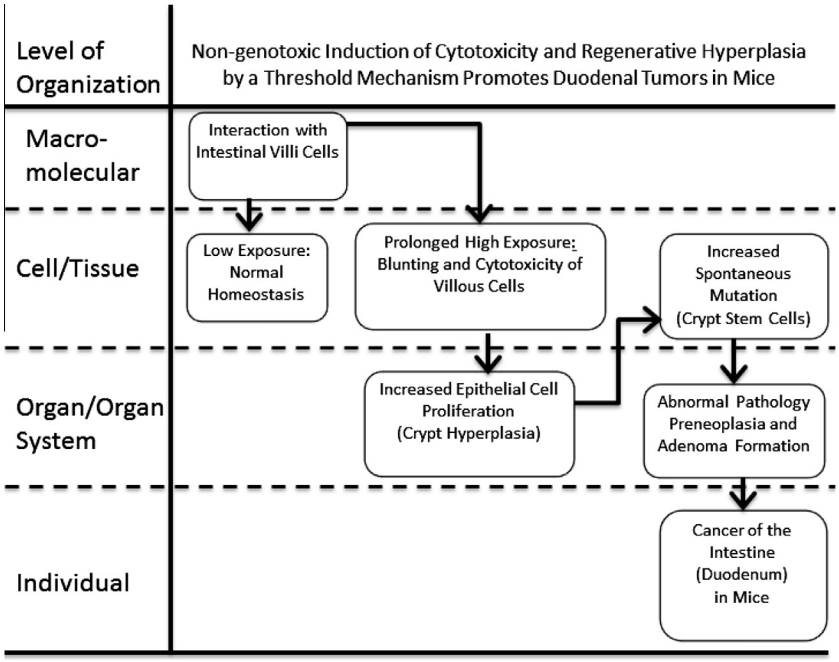

Fig. A5. Depiction of the AOP of induction of cytotoxicity and regenerative hyperplasia by oral CrVI leading to duodenal tumors in mice.

caused cytoxicity and tumorigenesis in the mouse. Furthermore, neither crypt hyperplasia nor intestinal tumors were seen in the NTP bioassay in rats. The evidence is determined to be less than strong because direct evidence such as from reversibility studies is not available at this time for CrVI. However, stop/reversibility studies with captan have shown that the tumor promotion responses are reversible, leading EPA to accept that "there is a strong causal association (dose–response, temporality) indicating that tumor formation is secondary to cytotoxicity and hyperplasia and that the latter is a KE in the sequential cascade of events leading to cancer" (USEPA, 2004).

A.5.3. Empirical support

The empirical support for each KE/KER is summarized below. Overall, the empirical support for the AOP is judged to be moderate to high (strong). As noted below, additional studies could provide greater knowledge that may bolster the WoE for a number of KEs/ KERs in this AOP.

A.5.3.1. Empirical support of pre-MIE events

The pre-MIE events for this case example AOP include: (1) reduction in the stomach and upper GI tract to CrIII; (2) absorption of any CrVI that is not reduced in the GI tract through sulfate and phosphate transporters by enterocytes (the columnar epithelial cells that form the intestinal villi). Note – CrIII is much less bioavailable than CrVI, thus extracellular reduction of CrVI to CrIII in the stomach is protective against the toxic and carcinogenic effects of oral exposure to CrVI (De Flora et al., 1997; De Flora, 2000; Proctor et al., 2012; Schlosser and Sasso, 2014). In terms of empirical support, a dose-dependent increase of chromium in duodenal epithelium has been observed. Although CrVI cannot be speciated in tissues, a significant increase in intestinal tissue chromium concentration is observed for concentrations greater than 5 mg/L CrVI (Thompson et al., 2013).

A.5.3.2. Empirical support of KER1 (doses exceeding the gastric reducing capacity lead to cytotoxicity in intestinal villi)

At sufficiently high oral doses of CrVI (e.g. doses that exceed the gastric reducing capacity), cytotoxicity is induced in the villous cells of the small intestine through contact with CrVI exiting the pylorus. This is observed and evaluated histopathologically as cytoplasmic vacuolization and blunting of the villi. A clear dose-threshold has been observed for these cytotoxic effects (Thompson et al., 2011). The empirical support of this KER is high (strong).

- Dose–response: villous cytotoxicity, measured by cytoplasmic vacuolization and villous atrophy has been reported to occur in a dose dependent fashion at doses which reduce/deplete capacity to reduce CrVI (Thompson et al., 2013).
- Temporal concordance: At 8 days, cytotoxicity was observed at 170 and 520 mg/L Sodium dichromate dihydrate (SDD). At 91 days, cytotoxicity was observed at 60 mg/L or greater, and at 2 years at 14.3 mg/L or greater.
- Incidence: neither NTP (2007) nor NTP (2008) report the incidence of cytotoxicity. Thompson et al. (2011) reported that cytoplasmic vacuolization in the duodenal villi was the most sensitive end point occurring at 170 and 60 mg/L SDD at days 8 and 91, respectively; while atrophy of the villi and crypt hyperplasia were first evident at higher concentrations (i.e. 520 and 170 mg/L on days 8 and 91, respectively).

A.5.3.3. Empirical support of KER2 (sustained cytotoxicity triggers crypt cell proliferation)

Hyperplasia occurs when stem cells in the crypts experience a prolonged stimulus to generate new epithelial cells. This is observed and evaluated histopathologically (Thompson et al., 2011, 2012, 2013). The empirical support of this relationship is high (strong).

- Species specificity: diffuse hyperplasia occurred in the duodenum of mice at all SDD concentrations examined in the 2-year bioassay, while there was no evidence of diffuse hyperplasia in the rat duodenum at any dose level.

Table A.5.3.5
Dose-response, temporal and incidence concordance of histopathological lesions in mice treated orally with CRVI.[a]

Females (dose) mg/kg-day	0	14	57	172	519
Diffuse hyperplasia	0%	32%	70%	62%	84%
Adenoma	0%	0%	4%	26%	24%
Carcinoma	0%	0%	0%	2%	12%
Males (dose) mg/kg-day	0	14	57	172	519
Diffuse hyperplasia	22%	36%	84%	84%	64%
Adenoma	2%	0%	2%	10%	30%
Carcinoma	0%	0%	0%	4%	6%

[a] From NTP (2008) and Thompson et al. (2011).

- Dose-response: the dose-response, temporal and incidence relationships between cytoplasmic vacuolization in villi (a marker of cytotoxicity) and crypt hyperplasia have been published by Thompson et al. (2013):
  - At Day 8: at 500 mg/L 100% cytoplasmic vacuolization and 60% crypt hyperplasia; at 170 mg/L 60% cytoplasmic vacuolization and 0% crypt hyperplasia; at 60 mg/L 0% cytoplasmic vacuolization and 0% crypt hyperplasia.
  - At Day 91: at 500 mg/L 70% cytoplasmic vacuolization and 90% crypt hyperplasia; at 170 mg/L 100% cytoplasmic vacuolization and 90% crypt hyperplasia; at 60 mg/L 50% cytoplasmic vacuolization and 0% crypt hyperplasia.

A.5.3.4. Empirical support of KER3 (crypt hyperplasia leads to an increased probability of spontaneous mutations in cells within the crypt)

The KE is postulated as sustained increase in proliferation, a more rapid cell cycle leading to increases in the probability of miss-repair of DNA and the fixation of spontaneous mutations in daughter cells. The exact relationship between the crypt hyperplasia and the occurrence of crypt cell mutations is not exactly known. The increase in stem cell population and number of cells transiting and dividing as they move up the villus likely leads to increased spontaneous mutation. Potential contributors to the proliferative environment may be the oxidative damage caused by CrVI entering the enterocytes resulting in anti-oxidant responses and infiltration of immune cells (NTP, 2008; Thompson et al., 2011, 2012, 2013). The empirical support of this relationship is moderate.

- Methodologies to measure spontaneous mutations are lacking, so the empirical evidence of this relationship is indirect.
- The evidence for a direct mutagenic mode of action of CrVI on duodenal epithelial cells is negative. O'Brien et al. (2013) reported no treatment-related effect Kras codon 12 even at the high CrVI doses that were carcinogenic in the 2-year bioassay and that increased crypt proliferation after 7 or 90 days of exposure. Furthermore, toxicogenomic data from mice exposed to CrVI for 90 days did not indicate changes in Apc gene expression.

A.5.3.5. Empirical support of KER4 (sustained increase in spontaneous mutations and sustained proliferative stimulus leads to formation of benign and malignant tumors)

An increase in spontaneous mutations coupled with sustained proliferative stimulus over a significant portion of the lifespan of the mouse provides a selective growth advantage to spontaneous initiated intestinal cells. The KE is formation of benign and malignant tumors due to clonal expansion and selection of cells that have lost normal growth regulatory controls. The empirical support of this relationship is high (strong). Again, the evidence is indirect, because empirical evidence of generation of spontaneous mutations is lacking.

- Dose-response: the dose-response, temporal and incidence relationships between hyperplasia in the duodenum in mice and adenomas and carcinomas of the duodenum are summarized in Table A.5.3.5.

References

Abe, R., Toyota, K., Miyakawa, H., Watanabe, H., Oka, T., Miyagawa, S., Nishide, H., Uchiyama, I., Tollefsen, E., Iguchi, T., Tatarazako, N., 2015. Diofenolan induces male offspring production through binding to the juvenile hormone receptor in Daphnia magna. Aquat. Toxicol. Aquat Toxicol. 159, 44–51. http://dx.doi.org/10.1016/j.aquatox.2014.11.015.

Abou Donia, M.B., Lapadula, D.M., 1990. Mechanisms of organophosphorus ester induced delayed neurotoxicity: type I and II. Ann. Rev. Pharmacol. Toxicol. 30, 405–440.

Alsharif, N.Z., Lawson, T., Stohs, S.J., 1994. Oxidative stress induced by 2,3,7,8-tetra chlorodibenzo-p-dioxin is mediated by the aryl hydrocarbon (Ah) receptor complex. Toxicology 6:92 (1-3), 39–51.

Bal-Price, A., Crofton, K.M., Sachana, M., Shafer, T.J., Behl, M., Forsby, A., Hargreaves, A., Landesmann, B., Lein, P.J., Louisse, J., Monnet-Tschudi, F., Paini, A., Rolaki, A., Schrattenholz, A., Suñol, C., van Thriel, C., Whelan, M., Fritsche, E., 2015. Workshop report: putative adverse outcome pathways (AOP) relevant to neurotoxicity. Crit. Rev. Toxicol. 45 (1), 83–91. http://dx.doi.org/10.3109/10408444.2014.981331.

Beischlag, T.V., Luis Morales, J., Hollingshead, B.D., Perdew, G.H., 2008. The aryl hydrocarbon receptor complex and the control of gene expression. Crit. Rev. Eukaryot. Gene Expr. 18, 207–250.

Boobis, A.R., Daston, G.P., Preston, R.J., Olin, S.S., 2009. Application of key events analysis to chemical carcinogens and noncarcinogens. Crit. Rev. Food. Sci. Nutr. 49, 690–707.

Casarett, L.J., Doull, J., 1922; Klaassen, C.D., Amdur, M.O., Casarett and Doull's Toxicology: the basic science of poisons, third ed., 1986. Curtis D. Klaassen, Mary O. Amdur, John Doull (Eds.). Macmillan, New York.

Case, R.M., Eisner, D., Gurney, A., Jones, O., Muallem, S., Verkhratsky, A., 2007. Evolution of calcium homeostasis: from birth of the first cell to an omnipresent signalling system. Cell Calcium 42, 345–350.

Cheng, L., Lai, M.-D., 2003. Aberrant crypt foci as microscopic precursors of colorectal cancer, World. J. Gastroenterol. 9, 2642–2649.

Chia, L.A., Kuo, C.J., 2010. The intestinal stem cell. Prog. Mol. Biol. Transl. Sci. 96, 157–173.

Choudhary, S., Gill, K.D., 2001. Protective effect of nimodipine on dichlorvos-induced delayed neurotoxicity in rat brain. Biochem. Pharmacol. 62, 1265–1272.

Cohen, S.M., Gordon, E.B., Singh, P., Arce, G.T., Nyska, A., 2010. Carcinogenic mode of action of folpet in mice and evaluation of its relevance to humans, Crit. Rev. Toxicol. 40, 531–545.

Cohen, S.M., Arnold, L.L., 2011. Chemical carcinogenesis. Toxicol. Sci. 120 (Suppl 1), S76–S92.

Costa, C., Catamnia, S., De Pasquale, R., Stancaenlli, R., Scribano, G.M., Melchini, A., 2010. Exposure of human skin to benzo(a)pyrene: role of CYP1A and aryl hydrocarbon receptor in oxidative stress generation. Toxicology 271, 83–86.

Curtis, L.R., Garzon, G.B., Arkoosh, M., Collier, T., Myers, M.S., Buzitis, J., Hahn, M.E., 2011. Reduced cytochrome P4501A activity and recovery from oxidative stress during subchronic benzo(a)pyrene and benzo(e)pyrene treatment of rainbow trout. Toxicol. Appl. Pharmacol. 254, 1–7.

Dalton, T.P., Puga, A., Shertzer, H.G., 2002. Induction of cellular oxidative stress by aryl hydrocarbon receptor activation. Chem-Biol. Interact. 141, 77–95.

De Flora, S., Camoirano, A., Bagnasco, M., Bennicelli, C., Corbett, G.E., Kerger, B.D., 1997. Estimates of the chromium(VI) reducing capacity in human body compartments as a mechanism for attenuating its potential toxicity and carcinogenicity. Carcinogenesis 18, 531–537.

De Flora, S., 2000. Threshold mechanisms and site specificity in chromium(VI) carcinogenesis. Carcinogenesis 21, 533–541.

Deroo, B.J., Korach, K.S., 2006. Estrogen receptors and human disease. J. Clin. Invest. 116, 561–570.

Doherty, J.D., 2006. Screening pesticides for neuropathogenicity. J. Biomed. Biotechnol. 2006, 70414.

EDSP, 2009. Endocrine Disruptor Screening Program Test Guidelines OPPTS 890.1300: Estrogen Receptor Transcriptional Activation (Human Cell Line (HeLa-9903)). Available at: <http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series890.htm>.

Emerick, G.L., DeOliveira, G.H., dos Santos, A.C., Ehrich, M., 2012. Mechanisms for consideration for intervention in the development of organophosphorus-induced delayed neuropathy. Chem-Biol. Interact. 199, 177–184.

Gielen, S., 2005. Estrogen Receptor Signaling in the Endometrium: pathway analysis of agonists and antagonists. http://repub.eur.nl/pub/7179/051102_Gielen-SCJP.pdf.

Gillette, J.R., Brodie, B.B., LaDu, B.N., 1957. The oxidation of drugs by liver microsomes. J. Pharmacol. Exp. Ther. 119, 532–540.

Giorgi, C., Romagnoli, A., Pinton, P., Rizzuto, R., 2008. Ca^{2+} signaling, mitochondria and cell death. Curr. Mol. Med. 8, 119–130.

Gordon, E., Cohen, S.M., Singh, P., 2012. Folpet-induced short term cytotoxic and proliferative changes in the mouse duodenum. Toxicol. Mech. Methods 22, 54–59.
Hargreaves, A.J., 2012. Neurodegeneration induced by organophosphorous compounds. Adv. Exp. Med. Biol. 724, 189–204.
Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., Tujague, M., Ström, A., Treuter, E., Warner, M., Gustafsson, J.A., 2007. Estrogen receptors: how do they signal and what are their targets. Physiol. Rev. 87, 905–931.
Johnson, M.K., 1990. Organophosphates and delayed neuropathy – is NTE alive and well? Toxicol. Appl. Pharmacol. 102, 385–389.
Johnson, L.R., 2013. Gastrointestinal Physiology. Elsevier Health Sciences.
Kalthoff, S., Ehmer, U., Manns, M.P., Strrassburg, C.P., 2010. Interaction between oxidative stress sensor nrf2 and xenobiotic-activated aryl hydrocarbon receptor in the regulation of the human phase II detoxifying UDP-glucuronosyltransferase 1A10. J. Biol. Chem. 26, 5993–6002.
Kato, Y., Kobayashi, K., Oda, S., Tatarazako, N., Watanabe, H., Iguchi, T., 2010. Sequence divergence and expression of a transformer gene in the branchiopod crustacean, Daphnia magna. Genomics 95, 160–165.
Kato, Y., Kobayashi, K., Watanabe, H., Iguchi, T., 2011. Environmental sex determination in the branchipod crustacean Daphnia magna: deep conservation of doublesex gene in the sex-determining pathway. PLoS Genet. 7 (3), e1001345.
Kopf, P.G., Scott, J.A., Agbor, L.N., Boberg, J.R., Elased, K.M., Huwe, J.K., Walker, M.K., 2010. Cytochrome P4501A is required for vascular dysfunction and hypertension induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin. Toxicol. Sci. 117, 537–546.
Kopf, P.G., Walker, M.K., 2010. 2,3,7,8-tetrachlorodibenzo-p-dioxinincreases reactive oxygen species production in human endothelial cells via induction of cytochrome P4501A1. Toxicol. Appl. Pharmacol. 245, 91–99.
Kuiper, G.G., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., van der Saag, P.T., van der Burg, B., Gustafsson, J.A., 1998. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinology 139 (10), 4252–4263.
Lapadula, D.M., Patton, S.E., Campbell, G.A., Abou-Donia, M.B., 1985. Characterisation of delayed neurotoxicity in the mouse following chronic oral administration of tri-o-cresyl phosphate. Toxicol. Appl. Pharmacol. 79, 83–90.
Latchoumycandane, C., Chitra, K.C., Mathur, P.P., 2002. Induction of oxidative stress in rat epididymal sperm after exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Arch Toxicol. 76, 113–118.
LeBlanc, G.A., Wang, Y.H., Holmes, C.N., Kwon, G., Medlock, E.K., 2013. A transgenerational endocrine signaling pathway in Crustacea. PLoS One 8 (4), e61715.
Legler, J., van den Brink, C.E., Brouwer, A., Murk, A.J., van der Saag, P.T., Vethaak, A.D., van der Burg, B., 1999. Development of a stably transfected estrogen receptor-mediated luciferase reporter gene assay in the human T47D breast cancer cell line. Toxicol. Sci. 48 (1), 55–66.
Lotti, M., Moretto, A., 2005. Organophosphate-induced delayed polyneuropathy. Toxicol. Rev. 24, 37–49.
Lu, H., Cui, W., Klaassen, C.D., 2011. Nrf2 protects against 2,3,7,8-tetrachlorodiben zo-p-dioxin-induced oxidative stress injury and steatohepatitis. Toxicol. Appl. Pharmacol. 256, 122–135.
McLachlan, J.A., 2001. Environmental signaling: what embryos and evolution teach us about endocrine disrupting chemicals. Endocr Rev. 22 (3), 319–341.
Miyakawa, H., Toyota, K., Hirakawa, I., Ogino, Y., Miyagawa, S., Oda, S., Tatarazako, N., Miura, T., Colbourne, J.K., Iguchi, T., 2013. A mutation in the methoprene tolerant alters juvenile hormone response in insects and crustaceans. Nat. Commun. 4, 1856.
Moggs, J.G., Tinwell, H., Spurway, T., Chang, H.S., Pate, I., Lim, F.L., Moore, D.J., Soames, A., Stuckey, R., Currie, R., Zhu, T., Kimber, I., Ashby, J., Orphanides, G., 2004. Phenotypic anchoring of gene expression changes during estrogen-induced uterine growth. Environ. Health Perspect. 112, 1589–1606.
Moser, M., Stempfl, T., Li, Y., Glynn, P., Büttner, R., Kretzschmar, D., 2000. Cloning and expression of the murine sws/NTE gene. Mech. Dev. 90, 279–282.
Mou, D.L., Wang, Y.P., Song, J.F., Rao, Z.R., Duan, L., Ju, G., 2006. Triorthocresyl phosphate-induced neuronal losses in lumbar spinal cord of hens–an immunohistochemistry and ultrastructure study. Int. J. Neurosci. 116, 1303–1316.
National Toxicology Program (NTP), 2007. NTP. Technical report on the toxicity studies of sodium dichromate dihydrate (CAS No. 7789-12-0). Administered in drinking water to male and female F344/N rats and B6C3F1 mice and male BALB/c and am3-C57BL/6 mice.
National Toxicology Program (NTP), 2008. NTP toxicology and carcinogenesis studies of sodium dichromate dihydrate (CAS No. 7789-12-0) in F344/N Rats and B6C3F1 Mice (Drinking Water Studies), Natl. Toxicol. Program Tech. Rep. Ser. 1–192.
Nebert, D.W., Roe, A.L., Dieter, M.Z., Solis, W.A., Yang, Y., Dalton, T.P., 2000. Role of aromatic hydrocarbon receptor and Ah gene battery in the oxidative stress response, cell cycle control and apoptosis. Biochem. Pharmacol. 59, 65–85.
O'Brien, T.J., Ding, H., Suh, M., Thompson, C.M., Parsons, B.L., Harris, M.A., Winkelman, W.A., Wolf, J.C., Hixon, J.G., Schwartz, A.M., Myers, M.B., Haws, L.C., Proctor, D.M., 2013. Assessment of K-Ras mutant frequency and micronucleus incidence in the mouse duodenum following 90-days of exposure to Cr(VI) in drinking water. Mutat. Res. 754, 15–21.
Oda, S., Tatarazako, N., Watanabe, H., Morita, M., Iguchi, T., 2005. Production of male neonates in four cladoceran species exposed to a juvenile hormone analog, fenoxycarb. Chemosphere 60, 74–78.
OECD, 2009. Test No. 455: The Stably Transfected Human Estrogen Receptor-alpha Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals. Availlable at: http://www.oecd-ilibrary.org/environment/test-no-455-the-stably-transfected-human-estrogen-receptor-alpha-transcriptional-activation-assay-for-detection-of-estrogenic-agonist-activity-of-chemicals_9789264076372-en;jsessionid=18jfavle748nr.x-oecd-live-03.
OECD, 2012. Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists. Available at: <http://www.oecd-ilibrary.org/environment/test-no-455-performance-based-test-guideline-for-stably-transfected-transactivation-in-vitro-assays-to-detect-estrogen-receptor-agonists_9789264185388-en>.
Olmstead, A.W., LeBlanc, G.A., 2003. Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean Daphnia magna. Environ. Health Perspect. 111, 919–924.
Palace, V.P., Kalverkamp, J.F., Lockhart, W.L., Metner, D.A., Muir, D.C.G., Brown, S.B., 1996. Mixed function oxidase enzyme activity and oxidative stress in lake trout (*Salvelinus namaycush*) exposed to 3,3′,4,4′,5-pentachlorobiphenyl (PCB-126). Environ. Toxicol. Chem. 15, 955–960.
Paris, M., Pettersson, K., Schubert, M., Bertrand, S., Pongratz, I., Escriva, H., Laudet, V., 2008. An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. BMC Evol. Biol. 8, 219.
Pitot, H.C., Hikita, H., Dragan, Y., Sargent, L., Haas, M., 2000. Review article: the stages of gastrointestinal carcinogenesis—application of rodent models to human disease. Aliment. Pharmacol. Ther. 14 (Suppl 1), 153–160.
Proctor, D.M., Suh, M., Aylward, L.L., Kirman, C.R., Harris, M.A., Thompson, C.M., Gürleyük, H., Gerads, R., Haws, L.C., Hays, S.M., 2012. Hexavalent chromium reduction kinetics in rodent stomach contents. Chemosphere 89, 487–493.
Reichard, J.F., Dalton, T.P., Shertzer, H.G., Puga, A., 2006. Induction of oxidative stress responses by dioxin and other ligands of the aryl hydrocarbon receptor. Dose Response 3, 306–331.
Richardson, R.J., Hein, N.D., Wijeyesakere, S.J., Fink, J.K., Makhaeva, G.F., 2013. Neuropathy target esterase (NTE): overview and future. Chem. Biol. Interact. 203, 238–244.
Schlezinger, J.J., White, R.D., Stegeman, J.J., 1999. Oxidative inactivation of cytochrome P-4501A (CYP1A) stimulated by 3,3′,4,4′-tetrachlorobiphenyl: production of reactive oxygen by vertebrates CYP1As. Mol. Pharmacol. 56, 588–597.
Schlezinger, J.J., Stegeman, J.J., 2001. Induction and suppression of cytochrome P4501A by 3,3′,4,4′,5-pentachlorobiphenyl and its relationship to oxidative stress in the marine fish scup (Stenotomus chyrsops). Aquat. Toxicol. 52, 101–115.
Schlosser, P.M., Sasso, A.F., 2014. A revised model of ex-vivo reduction of hexavalent chromium in human and rodent gastric juices. Toxicol. Appl. Pharmacol. 280, 352–361.
Senft, A.P., Dalton, T.P., Nebert, D.W., Gender, M.B., Puga, A., Hutchinson, R.J., Kerzee, J.K., Uno, S., Shertzer, H.G., 2002. Mitochondrial reactive oxygen production is dependent on the aromatic hydrocarbon receptor. Free Rad. Biol. Med. 33, 1268–1278.
Shaker, A., Rubin, D.C., 2010. Intestinal stem cells and epithelial-mesenchymal interactions in the crypt and stem cell niche. Transl. Res. 156, 180–187.
Shertzer, H.G., Nebert, D.W., Puga, A., Ary, M., Sonntag, D., Dixon, K., Robinson, L.J., Cianciolo, E., Dalton, T.P., 1998. Dioxin causes a sustained oxidative stress response in the mouse. Biochem. Biophys. Res. Commun. 253, 44–48.
Slezak, B.P., Hatch, G.E., DeVito, M.J., Diliberto, J.J., Slade, R., Crissman, K., Hassoun, E., Birnbaum, L.S., 2000. Oxidative stress in female B6C3F1 mice following acute and subchronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Toxicol. Sci. 54, 390–398.
Sonneveld, E., Riteco, J.A., Jansen, H.J., Pieterse, B., Brouwer, A., Schoonen, W.G., van der Burg, B., 2006. Comparison of in vitro and in vivo screening models for androgenic and estrogenic activities. Toxicol. Sci. 89 (1), 173–187.
Takahashi, M., Wakabayashi, K., 2004. Gene mutations and altered gene expression in azoxymethane-induced colon carcinogenesis in rodents. Cancer Sci. 95, 475–480.
Tatarazako, N., Oda, S., Watanabe, H., Morita, M., Iguchi, T., 2003. Juvenile hormone agonists affect the occurrence of male Daphnia. Chemosphere 53, 827–833.
Thompson, C.M., Proctor, D.M., Haws, L.C., Hébert, C.D., Grimes, S.D., Shertzer, H.G., Kopec, A.K., Hixon, J.G., Zacharewski, T.R., Harris, M.A., 2011. Investigation of the mode of action underlying the tumorigenic response induced in B6C3F1 mice exposed orally to hexavalent chromium. Toxicol. Sci. 123, 58–70.
Thompson, C.M., Proctor, D.M., Suh, M., Haws, L.C., Hébert, C.D., Mann, J.F., Shertzer, H.G., Hixon, J.G., Harris, M.A., 2012. Comparison of the effects of hexavalent chromium in the alimentary canal of F344 rats and B6C3F1 mice following exposure in drinking water: implications for carcinogenic modes of action. Toxicol. Sci. 125, 79–90.
Thompson, C.M., Proctor, D.M., Suh, M., Haws, L.C., Kirman, C.R., Harris, M.A., 2013. Assessment of the mode of action underlying development of rodent small intestinal tumors following oral exposure to hexavalent chromium and relevance to humans. Crit. Rev. Toxicol. 43, 244–274.
Thompson, C.M., Seiter, J., Chappell, M.A., Tappero, R.V., Proctor, D.M., Suh, M., Wolf, J.C., Haws, L.C., Vitale, R., Mittal, L., Kirman, C.R., Hays, S.M., Harris, M.A., 2015. Synchrotron-based imaging of chromium and γ-H2AX immunostaining in the duodenum following repeated exposure to Cr(VI) in drinking water. Toxicol. Sci. 143 (1), 16–25.
Thornton, J.W., 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc. Natl. Acad. Sci. U. S. A. 98, 5671–5676.

Toyota, K., Kato, Y., Sato, M., Sugiura, N., Miyagawa, S., Miyakawa, H., Watanabe, H., Oda, S., Ogino, Y., Hiruta, C., Mizutani, T., Tatarazako, N., Paland, S., Jackson, C., Colbourne, J.K., Iguchi, T., 2013. Molecular cloning of doublesex genes of four cladocera (water flea) species. BMC Genom. 14, 239.

Tremblay, A., Tremblay, G.B., Labrie, C., Labrie, F., Giguère, V., 1998. EM-800, a novel antiestrogen, acts as a pure antagonist of the transcriptional functions of estrogen receptors alpha and beta. Endocrinology 139 (1), 111–118.

Tsuji, G., Takahara, M., Uchi, H., Takeuchi, S., Mitoma, C., Moroi, Y., Furue, M., 2011. An environmental contaminant, benzo(a)pyrene induces oxidative stress-mediated interleukin-8 production in human keratinocytes via the aryl hydrocarbon receptor signalling pathway. J. Dermatol. Sci. 62, 42–49.

US EPA, 2004. Captan; cancer reclassification; amendment of reregistration eligibility decision; notice of availability. Fed. Regist. pp 68357–68360. http://www.gpo.gov/fdsys/pkg/FR-2004-11-24/pdf/04-26083.pdf.

Veronesi, B., Padilla, S., Blackmon, K., Pope, C., 1991. Murine susceptibility to organophosphorus-induced delayed neuropathy (OPIDN). Toxicol. Appl. Pharmacol. 108, 428–435.

Weiner, M.L., Jortner, B.S., 1999. Organophosphate-induced delayed neurotoxicity of triarylphosphates. Neurotoxicology 20, 653–673.

Zangar, R.C., Davydov, D.M., Verma, S., 2004. Mechanisms that regulate production of reactive oxygen species by cytochome P450. Toxicol. Appl. Pharacol. 199, 316–331.